(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 November 2001 (01.11.2001)

PCT

(10) International Publication Number
WO 01/81631 A1

(51) International Patent Classification[7]: C12Q 1/68, C12P 19/34

(21) International Application Number: PCT/US01/13136

(22) International Filing Date: 24 April 2001 (24.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/558,245 25 April 2000 (25.04.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/558,245 (CIP)
Filed on 25 April 2000 (25.04.2000)

(71) Applicant *(for all designated States except US)*: **DNA SCIENCES, INC.** [US/US]; 6540 Kaiser Drive, Fremont, CA 94555-3613 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **XU, Hua** [CN/US]; 1567 Samedra Street, Sunnyvale, CA 94087 (US). **MATHIES, Richard, A.** [US/US]; 93 Danefield Place, Moraga, CA 94556 (US).

(74) Agents: **AUSENHUS, Scott, L.** et al.; Townsend and Townsend and Crew LLP, 1200 Seventeenth Street, Suite 2700, Denver, CO 80202 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/81631 A1

(54) Title: DETECTION OF NUCLEOTIDE SEQUENCE VARIATIONS VIA THE PROOFREADING ACTIVITY OF POLYMERASES

(57) Abstract: The present invention provides methods and kits for determining the identity of a nucleotide at a variant site of a nucleic acid of interest, including point mutations and single nucleotide polymorphisms (SNPs). The methods utilize the proofreading activity of certain DNA polymerases in conjunction with selected detection primers to generate labeled and unlabeled extension products that are characteristic of the nucleotide at the variant site in the target nucleic acid. The methods are designed so that if the detection primer is complementary to the nucleotide at the variant site of the target nucleic acid, label from the detection primer is retained in the product. If, however, the primer is not complementary to the nucleotide at the variant site, then label from the detection primer is not retained in the product. In addition to their utility in detecting and analyzing point mutations and SNPs, the methods and kits of th4e invention have utility in a variety of other applications in which specific nucleic acid - information is of value, including detection of pathogens, resolution of paternity disputes, prenatal testing and forensic analysis.

DETECTION OF NUCLEOTIDE SEQUENCE VARIATIONS VIA THE PROOFREADING ACTIVITY OF POLYMERASES

CROSS-REFERENCES TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Patent Application No. 09/558,245, filed April 25, 2000, which is incorporated herein in its entirety for all purposes.

FIELD OF INVENTION

The present invention relates to the field of molecular genetics, particularly the identification and detection of particular nucleotide sequences.

BACKGROUND OF THE INVENTION

Many diseases linked to genome modifications, either of the host organism or of infectious organisms, are often the consequence of a change in a small number of nucleotides, frequently involving a change in a single nucleotide. Such single nucleotide changes are referred to as single nucleotide polymorphisms or simply SNPs, and the site at which the SNP occurs is typically referred to as a polymorphic site.

The ability to detect specific nucleotide alterations or mutations in DNA sequences is an important medical tool. Methods capable of identifying such alterations provide a means for diagnosing many common diseases that are associated with SNPs, such as diabetes, thalassemia, sickle-cell anemia, cystic fibrosis, and certain oncogenic mutations. Methods that can quickly identify such changes or mutations are also valuable in taking prophylactic measures, assessing the propensity for disease and in patient counseling and education. Additionally, such methods have value in a variety of non-medical applications, including the detection of microorganisms, resolving paternity disputes and in forensic analysis.

Various methods have been developed to obtain sequence information for variants sites. Such methods include allele-specific oligonucleotide hybridization, allele specific amplification, mini-sequencing methods, quantitative RT-PCR methods (*e.g.*, the so-called "TaqMan assays"), and various single base pair extension assays. However, frequently

these methods include one or more shortcomings such as high cost, relatively long analysis times, use of radiolabeled reagents, complexity and/or being poorly suited for multiplex analysis. For example, certain methods generate a positive result only for one of the possible nucleotides present at a variant site. With such methods, it is necessary to conduct additional analyses to ensure that the negative result is not the consequence of a failed experiment. Other methods do not retain label throughout the analysis, thus making such methods poorly suited for conducting multiplex analyses in which different labels can be utilized to interrogate multiple variant sites simultaneously.

SUMMARY OF THE INVENTION

The present invention provides methods and kits useful for detecting and identifying the nucleotide at a variant site of a target nucleic acid of interest. The methods and kits can be utilized in research, clinical and laboratory settings. In general, the methods selectively generate labeled and/or unlabeled extension products that are characteristic of the nucleotide at the variant site. The methods can be used in conducting genotyping analyses and can be performed in multiplexing formats. The methods find utility in diverse applications including, for example, analyzing point mutations and single nucleotide polymorphisms, detection of pathogens, resolution of paternity disputes, prenatal testing and forensic investigations.

More specifically, certain methods for analyzing a variant site in a target nucleic acid involve conducting a sequence-specific amplification which comprises amplifying the target nucleic acid by extending a first and second primer in the presence of a polymerase having a 3'-5' exonuclease activity. The first primer anneals to a segment that spans the variable site in a first strand of the target nucleic acid and bears a label on at least one nucleotide other than the 3' terminal nucleotide. The second primer is complementary to a segment of a second strand of the target nucleic acid which is complementary to the first strand. If the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, the first primer is extended to produce a product retaining the label. If, however, the first primer lacks complementarity to the base occupying the variable site in the first strand of the target nucleic acid, the 3'-5' exonuclease activity of the polymerase digests bases from the 3'-end of the first primer, thereby removing the label from the first primer; the polymerase then extends the first primer to produce a product lacking the

label. Product is then analyzed to determine whether the product retains the label, the presence or absence of label indicating the nucleotide at the variant site.

A variety of labeling modifications based on this general method can be utilized. For example, in one variation, the second primer is labeled with a second label that is distinguishable from the label borne by the first primer. In this variation, if the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, then product formed retains both labels (*i.e.*, the label from the first primer and the second label of the second primer); whereas, if the first primer lacks complementarity to the base occupying the variant site in the first stand of the target nucleic acid, then product formed bears only the second label. Products are then analyzed to determine whether the product bears the first and/or second label.

In another variation, the first primer bears two labels, a first label attached to a nucleotide other than the 3' terminal nucleotide and a second label attached 5' of the first label and at a nucleotide at which it cannot be removed during the proofreading process. If the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, then the product formed retains both labels. If, however, the first primer lacks complementarity to the base occupying the variant site on the first strand of the target nucleic acid, only the second label is retained. The product is then analyzed to detect product to determine if it retains the first and/or second label.

In yet another variation, the product being analyzed is duplex DNA and the label borne by the primer is a fluorescent label. In certain methods, the duplex DNA product is exposed to a fluorescent intercalation dye to introduce a plurality of dye molecules into labeled and/or unlabeled product (*i.e.*, product retaining the fluorescent label and/or product not retaining the fluorescent label) and then detecting fluorescence from the label and/or the intercalation dye. In some methods, the fluorescent dye is sufficiently close to the intercalation dye in the labeled product such that energy transfer between the fluorescent dye and the intercalation dye can occur. In some situations, the intercalation dye is excited and then fluorescence from the fluorescent dye detected.

Some methods are performed with multiple labeled primers selected to be complementary to each of the allelic forms of the target nucleic acid. Certain of these methods comprise conducting a sequence-specific amplification step, the amplification step comprising amplifying the target nucleic acid in a sample containing multiple copies of the

target nucleic acid by extending a plurality of primer pairs in the presence of a polymerase having a 3'-5' exonuclease activity, each primer pair comprising a first and second primer. Each of the first primers anneals to a segment that spans the variable site in a first strand of the target nucleic acid and bears a label on at least one nucleotide other than the 3' terminal nucleotide. The first primers are selected to be complementary to each of the allelic forms of the target nucleic acid such that the first primers in the different primer pairs are complementary to a different one of the plurality of bases that potentially occupy the variant site and bear different labels. The second primer of each pair is complementary to a segment of a second strand of the target nucleic acid. The first and second primers flank the variant site once hybridized to the target nucleic acid. For each primer pair, if the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, the first primer is extended to produce a product retaining the label. Consequently, the extension reactions generate a differentially labeled product for each of the allelic forms of the target nucleic acid present in the sample. The labeled product(s) are subsequently detected, the identity of the label indicating the allelic form(s) of the target nucleic acid present in the sample.

The invention also provides multiplexing methods in which multiple variant sites are analyzed together. Some of these methods involve conducting a sequence-specific amplification step, the amplification step comprising amplifying one or more target nucleic acids by extending a plurality of primer pairs in the presence of a polymerase that has a 3'-5' exonuclease activity, each primer pair comprising a first and second primer. Each first primer anneals to a segment that spans a variable site in a target nucleic acid and bears a label on at least one nucleotide other than the 3' terminal nucleotide, and each second primer is complementary to a segment of a second strand of the target nucleic acid, with the first and second primes flanking the variant site once hybridized to the target nucleic acid. Different primer pairs span different variant sites so that the first and second primers for each of the different primer pairs flank different variant sites. For each primer pair, if the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, the first primer is extended to produce a product retaining the label. However, if the first primer lacks complementarity to the base occupying the variable site in the first strand of the target nucleic acid, the 3'-5' exonuclease activity of the polymerase digests bases from the 3'-end of the first primer, thereby removing the label from the first primer; the

polymerase then extends the primer to produce a product lacking the label. Thus, a plurality of products are formed from the different primer pairs. The products are then analyzed, the analyzing step comprising determining whether the products retain label to identify the nucleotide at the different variant sites.

While the foregoing methods utilize two primers to amplify both strands of the target nucleic acid, linear amplification methods similar to the methods just described can be performed utilizing a single labeled primer. In this instance, the general method typically involves conducting a sequence-specific primer extension step, the extension step comprising extending at least one primer that anneals to a segment of the target nucleic acid that spans the variable site and bears a label on at least one nucleotide other than the 3'-terminal nucleotide in the presence of a polymerase having a 3'-5' exonuclease activity. If the primer is complementary to the base occupying the variable site in the target nucleic acid, the primer is extended to produce a product retaining the label. Whereas, if the primer lacks complementarity to the base occupying the variable site in the target nucleic acid, the 3'-5' exonuclease activity of the polymerase digests bases from the 3'-end of the primer, thereby removing the label from the primer, and the polymerase then extends the primer to produce a product lacking the label. Product is then analyzed by determining whether the product bears the label to identify the nucleotide at the variant site. The methods are typically conducted under thermal cycling conditions to amplify the labeled and/or unlabeled product.

As with the two-primer amplification methods described above, a variety of different labeling schemes can be used in variations of the general method. For instance, the primer can bear multiple labels, including a first label attached at a nucleotide other than the 3' terminal nucleotide and a second label attached 5' of the first label and attached to a nucleotide from which it cannot be excised during the proofreading process.. In such methods, the primer bears a second label on a nucleotide other than the one to which the label is attached. If the primer is complementary to the base occupying the variant site of the target nucleic acid, then product formed retains both labels. If, however, the primer lacks complementarity to the base occupying the variant site on the target nucleic acid, only the second label is retained. The product is then analyzed to detect product that retains the label and/or second label.

In another variation, analyses are conducted with multiple labeled primers complementary to the different allelic forms of the target nucleic acid. Certain of these

methods involve conducting a sequence-specific amplification step, the amplification step comprising amplifying the target nucleic acid in a sample containing multiple copies of the target nucleic acid by extending a plurality of primers in the presence of a polymerase having a 3'-5' exonuclease activity, wherein each of the primers anneals to a segment that spans the variable site in the target nucleic acid and bears a label on at least one nucleotide other than the 3' terminal nucleotide. The primers are selected to be complementary to each of the allelic forms of the target nucleic acid such that different primers are complementary to a different one of the plurality of bases that potentially occupy the variant site and bear different labels. For each primer, if the primer is complementary to the base occupying the variant site in the target nucleic acid, the primer is extended to produce a product retaining the label. Hence, the extension reactions generate a differentially labeled product for each of the allelic forms of the target nucleic acid. The labeled product(s) are then detected, the identity of the label indicating the allelic form(s) of the target nucleic acid present in the sample.

The linear amplification methods can also be performed in a multiplexing format. Certain multiplexing methods include conducting a sequence-specific amplification that comprises amplifying one or more target nucleic acids by extending a plurality of primers in the presence of a polymerase having a 3'-5' exonuclease activity, wherein each primer has a different sequence and anneals to a segment that spans a different variable site in a target nucleic acid and bears a label on at least one nucleotide other than the 3' terminal nucleotide. For each of the different primers, if the primer is complementary to the base occupying the variant site in the target nucleic acid, the first primer is extended to produce a product retaining the label. If, in contrast, the primer lacks complementarity to the base occupying the variable site in the target nucleic acid, the 3'-5' exonuclease activity of the polymerase digests bases from the 3'-end of the first primer, thereby removing the label from the first primer, and wherein the first primer is thereafter extended by the polymerase to produce a product lacking the label. As a consequence of the extension reactions, a plurality of products are formed. These products are analyzed by determining whether the products retain label to identify the nucleotide at the different variant sites. Products formed from different variant sites can be distinguished by using differentially labeled primers and/or by differences in size, differentially labeled and/or different sized extension products being generated from different variant sites.

The invention also provides kits for conducting analyses such as those described above. Certain kits for conducting linear amplification of target nucleic acids include (i) one or more primers, each primer having a different sequence and capable of annealing to a segment that spans a variable site in a strand of a target nucleic acid and bearing a label or having an attachment moiety for attaching a label on at least one nucleotide other than the 3' terminal nucleotide, different primers bearing different labels or having different attachment moieties, and (ii) a polymerase having 3'-5' exonuclease activity. The primers in such kits can be complementary to different allelic forms of the target nucleic acid, wherein primers complementary to different allelic forms are differentially labeled. Alternatively, the primer can bear a second label attached to a nucleotide other than the one to which the label is attached.

Kits for conducting exponential amplification of target nucleic acids include forward and reverse primers. For example, certain kits comprise one or more primer pairs, each primer pair comprising a first and second primer, wherein (i) each first primer anneals to a segment that spans a variable site in a first strand of a target nucleic acid and bears a label or an attachment moiety for attaching a label on at least one nucleotide other than the 3' terminal nucleotide, first primers from different pairs having a different sequence and bearing different labels or having different attachment moieties, and (ii) each second primer is complementary to a segment of a second strand of the target nucleic acid downstream of the variant site. The kit also includes a polymerase having 3'-5' exonuclease activity.

Certain other kits include primers that have modified nucleotides that allow for the attachment of a label to the primer. For instance, some kits include: (i) one or more primers, each primer comprising a modified nucleotide for attachment of label and capable of annealing to a segment that spans the variable site in a strand of a target nucleic acid, the modified nucleotide being other than the 3' nucleotide, (ii) a label that can be attached to the first primer, and (iii) a polymerase having 3'-5' exonuclease activity.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A-1B illustrates the elements and general approach for determining the identity of a nucleotide at a variant site in a target nucleic acid according to certain methods of the invention. FIG. 1B is an exploded view of the region encircled in FIG. 1A.

FIG. 1C depicts certain steps in an example of one method of the invention using a single labeled detection primer.

FIG. 1D depicts certain steps in an example of one method of the invention using two labeled detection primers.

FIGS. 2A-2D relate to different aspects of a genotyping experiment for the polymorphism LPL (Lipoprotein Lipase Precursor) SNP-19 (T/G polymorphism at position 8393) (see, *e.g.*, Nickson, D.A. *et al, Nature Genetics* 19:233-240,1998). FIG. 2A lists the partial nucleotide sequence of LPL surrounding the LPL-19 SNP (as indicated by "k" to indicate the presence of T or G). Primer binding sites are underlined. FIG. 2B shows the detection primer utilized for the LPL-19 SNP "T" allele. The detection primer is complementary to the complement of the sequence shown in FIG. 2A. FIG. 2C shows agarose gel electrophoresis results indicating that PCR products of the expected size were formed for samples obtained from the different test subjects PD1 (a T homozygote), PD2 (a T/G heterozygote), PD3 (a T/T homozygote), PD5 (a T/G heterozygote) and PD8 (a G/G homozygote). FIG. 2D shows plots of fluorescence intensity versus time corresponding to the extension products formed for each of the test subjects using a primer having a T nucleotide at the variant site as separated by denaturing gel electrophoresis on a micro-channel plate (MCP).

FIGS. 3A-3D relate to different aspects of a genotyping experiment for the polymorphism LPL (Lipoprotein Lipase Precursor) SNP-9 (A/C polymorphism at position 5554). FIG. 3A lists a partial sequence of LPL encompassing the LPL-9 SNP site (as indicated by the "m" to indicate the presence of A or C). Primer binding sites are underlined. FIG. 3B provides the detection primer sequence for an artificial "T" allele at the LPL-9 SNP site. The detection primer is complementary to the complement of the sequence shown in FIG. 3A. FIG. 3C shows agarose gel electrophoresis results indicating successful amplification of a segment of the target nucleic acid containing the variant site for samples obtained from the different test subjects PD1 (A/A homozygote), PD3 (A/C heterozygote), PD8 (C/C homozygote) and a positive control sequence containing T at the variant site. FIG. 3D shows plots of fluorescence intensity versus time corresponding to the extension products formed for each of the test subjects and a control using a primer having a T nucleotide at the variant site as separated by denaturing gel electrophoresis on a MCP.

FIGS. 4A-4C depict additional results for the polymorphism LPL (Lipoprotein Lipase Precursor) SNP-9 (A/C polymorphism at position 5554). FIG. 4A lists the sequences and shows the label attachment site for the two detection primers (Primer-1 and Primer-2) for the LPL-9 SNP "T" allele used in the experiments. The nucleotide (dT) aligned with the polymorphic site was labeled with the dye ROX. The 3' terminus of the primers extended 1 (control Ctr-1 and sample PD3-1) or 2 (control Ctr-2 and sample PD3-2) nucleotides downstream from the polymorphic site. FIG. 4B shows agarose gel electrophoresis results indicating successful amplification of a segment of the target nucleic acid containing the variant site using Primer-1 and Primer-2. Lanes labeled No-1 and No-2 are negative controls in which template DNA was omitted from the PCR reactions. Samples from subject PD3 included A and C at the variant site. The control samples (Ctrl-1, Ctrl-2) contain a "T" residue at the SNP site. FIG. 4C shows plots of fluorescence intensity versus time corresponding to the extension products formed for each of the various trials utilizing Primer-1 and Primer-2 as separated by denaturing gel electrophoresis on a MCP.

FIG. 5 illustrates the general approach and steps for determining the identity of a nucleotide at a variant site in a target nucleic acid using one of the two-color proofreading assay systems in which amplification reactions are conducted with a labeled detection primer and a labeled reverse primer.

FIG. 6 illustrates an exemplary homogeneous proofreading assay utilizing fluorescent polarization to distinguish between labeled amplified product and labeled nucleotides in the reaction mixture.

FIG. 7A-7C relate to different aspects of a genotyping experiment for the polymorphism LPL (Lipoprotein Lipase Precursor) SNP-5 (A/T polymorphism at nucleotide 4343) (see, *e.g.*, Nickson, D.A. *et al, Nature Genetics* 19:233-240,1998). FIG. 7A lists the partial nucleotide sequence of LPL surrounding the LPL SNP-5 (as indicated by "W" to indicate the presence of A or T). Primer binding sites are underlined. FIG. 7B shows the detection primer utilized for the LPL SNP-5 "T" allele; this primer is complementary to the complement of the sequence shown in FIG. 7A. The ROX-labeled reverse primer is shown in FIG. 7C.

FIG. 8 shows genotyping results for the LPL-5 SNP utilizing a two-color proofreading assay such as illustrated in FIG. 5 with the R110-labeled detection primer and the ROX-labeled reverse primer shown in FIGS. 7B and 7C. The plots shown in FIG. 8 are

electropherograms (plots of fluorescence intensity versus time) showing the extension products formed for each of six test subjects as separated by nondenaturing gel electrophoresis on a MegaBACE (Amersham) utilizing a primer having a T nucleotide at the variant site (see FIG. 7B). Extension products formed from both the A and T alleles contain the ROX label (red; indicated by the letter R) from the labeled reverse primer . However, only extension products generated from the T allele contain the R110 label (blue; indicated by the letter B), as this label is removed via the proofreading activity of the polymerase for the A allele. The results show that the expected PCR products were formed for each sample obtained from the different test subjects PDR02 (an A homozygote), PDR05 (an A homozygote), PDR03 (an A/T heterozygote), PDR06 (an A/T heterozygote), PDR08 (a T/T homozygote) and HD1.10 (a T/T homozygote).

FIG. 9 is a plot of ΔFP-R110/ ΔFP-ROX for 57 different individuals that were genotyped at the LPL SNP-5 utilizing the homogeneous fluorescence polarization proofreading assay illustrated in FIG. 6 and the primers shown in FIGS. 7B and 7C. Thus, the values reflect a ratio of the change in fluorescence polarization (FP) for the T allele (since R110 is attached to the detection primer with T at the variant site) and the change in FP for signal from label on the reverse primer (ROX label attached to the reverse primer). As described with regard to FIG. 8, ROX label is incorporated into all amplification products regardless of the nucleotide at the polymorphic site; whereas, R110 appears only in extension product generated from the T allele. The results show that utilizing such an approach that the values for all but one individual of the 57 tested cluster into three distinct groups depending upon the actual genotype of the individual (i.e., A/A homozygote, A/T heterozygote and T/T homozygote). Experimental details and a further discussion of these results are provided in Example 2 infra.

DETAILED DESCRIPTION

I. Definitions

A "nucleic acid" is a deoxyribonucleotide or ribonucleotide polymer in either single or double-stranded form, including known analogs of natural nucleotides unless otherwise indicated.

A "polynucleotide" refers to a single - or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases.

An "oligonucleotide" is a single-stranded nucleic acid typically ranging in length from 2 to about 500 bases. Oligonucleotides are often synthetic but can also be produced from naturally occurring polynucleotides. Oligonucleotides can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphotriester method of Narang *et al.*, *Meth. Enzymol.* 68:90-99 (1979); the phosphodiester method of Brown *et al.*, *Meth. Enzymol.* 68:109-151 (1979); the diethylphosphoramidite method of Beaucage *et al.*, *Tetrahedron Lett.* 22:1859-1862 (1981); and the solid support method described in U.S. Patent No. 4,458,066.

A "primer" is a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (*i.e.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term "primer site" refers to the segment of the target DNA to which a primer hybridizes. The term "primer pair" means a set of primers including a 5' "upstream primer" that hybridizes with the 5' end of the DNA sequence to be amplified and a 3' "downstream primer" that hybridizes with the complement of the 3' end of the sequence to be amplified.

A primer that is "perfectly complementary" has a sequence fully complementary across the entire length of the primer and has no mismatches. The primer is typically perfectly complementary to a portion (subsequence) of a target sequence. A "mismatch" refers to a site at which the nucleotide in the primer and the nucleotide in the target nucleic acid with which it is aligned are not complementary.

The term "substantially complementary" means that a primer is not perfectly complementary to its target sequence; instead, the primer is only sufficiently complementary to hybridize selectively to its respective strand at the desired primer binding site.

A "variant site" broadly refers to a site wherein the identity of nucleotide at the site varies between nucleic acids that otherwise have similar sequences. For double-stranded nucleic acids, the variant site includes the variable nucleotide on one strand and the

complementary nucleotide on the other strand. A variant site can be the site of a single nucleotide polymorphism or the site of a somatic mutation, for example.

A "polymorphic marker" or "polymorphic site" is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wild-type form, whereas allelic forms occurring less frequently are referred to as mutant alleles. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms.

A "single nucleotide polymorphism" (SNP) occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations). A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

The term "naturally occurring" as applied to an object means that the object can be found in nature.

The term "subject" and "individual" are used interchangeably herein to refer to any type of organism, but most typically is used to refer to a human.

II. Overview

The present invention provides a variety of methods and kits for determining the identity of a nucleotide present at a variant site in a target nucleic acid. The methods are based, in part, upon the recognition that the proofreading activity of certain DNA polymerases can be used in conjunction with selected labeled or modified detection primers to generate labeled and unlabeled extension products that are characteristic of the nucleotide at the variant site in the target nucleic acid of interest. More specifically, the methods are

designed so that if the detection primer is complementary to the nucleotide at the variant site of the target nucleic acid, then the label from the detection primer is retained. If, however, the detection primer is not complementary to the nucleotide at the variant site, then the label from the detection primer is removed through the proofreading activity of the polymerase. The fact that an extension product of some type is formed regardless of the identity of the nucleotide at the variant site is important in that one can distinguish between the situation in which the primer is not complementary to the nucleotide at the variant site and a failed experiment in which an extension product is not produced. Even product in which label was removed during proofreading can be easily detected using various detection schemes (*e.g.*, by exposing the unlabeled product to an intercalation fluorescent dye or using a labeled reverse/second primer during amplification).

Because the methods generate labeled products, the methods can be conducted in multiplex formats in which the identity of multiple nucleotides at different variant sites is interrogated in the same reaction. Such methods can be performed by using different primers bearing different labels for each variant site and/or by using different primers bearing the same label but generating different sized products (*e.g.*, by the choice of second/reverse primer location). Thus, for example, the methods of the invention can be used to determine the identity of the nucleotide at multiple variant sites on a single nucleic acid template or at multiple variant sites on different nucleic acids. This multiplexing feature enables certain methods to achieve very high throughput.

The methods can be used in a number of different applications. For example, in the medical field, the methods of the invention can be used to determine which allele is present at a single nucleotide polymorphic (SNP) site or to detect mutations at a particular site. Because many diseases are associated with SNPs or mutations, the methods can be used in a variety of diagnostic, research and prognostic applications. In addition, for diploid subjects, the methods can be used to determine if the individual is homozygous or heterozygous for a particular allele at the variant site, *i.e.*, to determine the genotype of the individual. This is an important capability because individuals that are homozygous for an allele associated with a disease are at greater risk than individuals that are heterozygous or homozygous for the non-disease associated allele. Furthermore, individuals that are homozygous for an allele associated with a particular disease sometimes suffer the symptoms of the disease to a greater extent than heterozygotes. The ability of the methods to interrogate

particular sites also is useful for identification purposes, including for example, forensic and paternity cases. The methods also have utility in detecting the presence of nucleic acids from particular pathogens (*e.g.*, certain viruses, bacteria or fungi).

III. Determination of Nucleotide at Variant Site

A. General Description

As shown in FIG. 1A, certain methods are initiated by annealing a first primer (also called a detection primer or forward primer) to a first strand (*i.e.*, the template strand) of a target nucleic acid that includes a variant site. The particular target nucleic acid and primer sequences shown in FIG. 1A are for illustrative purposes only. Primers and target nucleic acids of differing sequences and lengths can be utilized as described further below. The term "target nucleic acid" refers to single- or double-stranded nucleic acids that include the variant site being interrogated. For double-stranded DNA, the variant site includes the base at the site being interrogated and the complementary nucleotide in the complementary strand (see FIG. 1A). As used herein, the term "template strand" refers to the strand to which the detection primer anneals. The 3' end of the detection primer when annealed to the template strand of the target nucleic acid spans the variant site in the template strand. As used herein, the phrase "spans the variant site" means that the primer when annealed to the template strand at least extends so as to anneal to the nucleotides that are immediately 5' and 3' of the variant site. Hence, the 3' end of the detection primer is aligned with a nucleotide in the template strand located 5' of the variant site in the template strand.

The detection primer includes a labeled nucleotide or a modified nucleotide that facilitates labeling of the first primer or is otherwise detectable. For simplicity of reference, such nucleotides are simply referred to herein as "labeled nucleotides" or "labeled bases." Thus, the term nucleotide when used to refer to the moiety to which the label is attached refers to both a naturally occurring nucleotide or modified nucleotides that facilitate label attachment within a primer. Similarly, references to labeled primers or primers that bear a label can mean primers which are attached to a label or a primer that includes a nucleotide such as a modified nucleotide that can be selectively labeled after extension reactions are completed. Likewise, reference to a labeled extension product refers both to extension products that actually include a label as well as extension products that include a

modified nucleotide capable of being labeled after the extension reaction or which has an inherently detectable property.

The labeled nucleotide is located in a segment of the detection primer that extends from the nucleotide aligned with the nucleotide at the variant site in the template strand to the nucleotide immediately adjacent the 3' terminal nucleotide of the primer (*i.e.*, the nucleotide just 5' of the 3' terminal nucleotide; see FIG. 1B). This segment is referred to as the "label attachment region." In some methods, the labeled nucleotide is aligned with the variable site. In other instances, if the nucleotide positions in the primer are numbered from the 3' end with the 3' terminal nucleotide being designated the first position base, the labeled nucleotide is at the second, third, fourth or fifth base position.

By avoiding labeling the 3' terminal nucleotide of the detection primer, the methods provide for more efficient extension reactions, as labels at the 3' terminal end can inhibit the extension reactions by inhibiting the polymerase. The label should not be attached to a nucleotide positioned 5' to the nucleotide in the detection primer that is aligned with the variant site in the template strand. This ensures that the proofreading activity of the polymerase removes the labeled or modified nucleotide in those instances in which the detection primer is not complementary to the nucleotide at the variant site of the target nucleic acid (see *infra*).

The annealed detection primers are then extended in reaction mixtures that include the necessary deoxyribonucleoside triphosphates (dATP, dTTP, dCTP and dGTP) and a polymerase having proofreading activity (*i.e.*, the ability to initially excise mismatched base pairs in the duplex formed between the primer and template strand before extending the primer) (see FIG. 1C). A variety of such polymerases are commercially available, including for example, Pfu (Stratagene), Tli (Promega), Vent (New England BioLabs), Deep Vent (New England BioLabs) and Pfx (Life Technologies).

As illustrated in FIG. 1C, if the nucleotide in the detection primer that is aligned with the variant site in the template strand of the target nucleic acid is complementary to the nucleotide at the variant site, the polymerase extends the detection primer to produce an extension product in which the labeled nucleotide is retained. If, however, the detection primer is not complementary with the nucleotide at the variant site in the template strand, then the polymerase initially digests the nucleotides from the 3'-end of the primer back to, and including, the site of the mismatch via its proofreading activity. In so doing, the

polymerase excises the labeled nucleotide. The polymerase subsequently extends the primer to produce an extension product in which the label is not retained. Hence, the formation of product that retains or lacks the label from the detection primer is characteristic of which nucleotide is present at the variant site.

FIG. 1C depicts a specific example for a T/C polymorphism. In the particular example shown in FIG. 1C, the detection primer includes the nucleotide T at the base that aligns with the variant site of the template strand. In the case of the T allelic form, the detection primer is complementary to the base occupying the variant site of the template strand (*i.e.*, the nucleotide A) and amplified product retains the label borne by the detection primer. For the C allelic form, however, the detection primer is not complementary with the base occupying the variant site of the template strand. Consequently, the label is not retained in the amplified product.

In certain other methods, the identity of all the polymorphic forms for a particular variant site can be determined simultaneously using multiple detection primers bearing different labels. FIG. 1D illustrates an example for a biallelic situation (T and C allelic forms present), in which the two allelic forms are interrogated using two detection primers bearing different labels. The steps performed are the same as those described for methods using a single detection primer. However, as can be seen in FIG. 1D, using differentially labeled detection primers for each allelic form, different labeled extension products are formed for each allelic form. Two unlabeled extension products are also generated, but the labeled extension products are the products typically detected because the label facilitates detection. However, as described more fully *infra*, labeled reverse primers can be used so that even extension products lacking label from the detection primer are nonetheless labeled by incorporation of label from the second primer.

The extension reactions can simply be linear amplification reactions in which a single primer (the detection primer) is extended to form extension product. The signal to noise ratio in such methods can be increased by repeating the annealing and extension reaction multiple times utilizing a thermostable polymerase and a thermocycler.

By utilizing exponential amplification reactions in the extension step, further improvement of the signal to noise ratio can be achieved. In such instances, a second primer (also called a reverse primer) that anneals to a segment of a second strand that is complementary to the template strand (*i.e.*, the complement of the template strand) is

included in the extension reactions (see FIG. 1A). The second primer anneals at a segment that is downstream of the first primer. Hence, the detection primer and second primer once annealed to their respective strands allow amplification of both strands of a duplex target nucleic acid.

The second primer can be used to increase the options for conducting analyses. For example, the choice of the sequence for the second primer governs where the primer binds and consequently the size of the product formed. Thus, as described more fully below in the multiplexing section, the choice of second primer can be used to generate extension products of differing size for different variant sites. Because the extension products can be size differentiated, extension products from multiple sites can be analyzed together (*e.g.*, by a size separation method such as electrophoresis). The second primer can also include a modified or labeled nucleotide to aid in detection or separation from other reactants. In one approach, the second primer carries a label that differs from the label attached to the detection primer. This allows product in which the detection label is removed during proofreading to still be easily detected via the second primer label (so long as double-stranded DNA is detected). By using a different label than that used in the detection primer, one can determine in a single reaction whether one or two allelic forms are present in a sample. Additionally, the second primer can include a modified moiety to facilitate separation of the amplification products from other reactants (*e.g.*, a moiety that allows for affinity separation; see the section on primers *infra*).

Methods for conducting the extension reactions (either linear or exponential amplifications) are well-known in the art (see, *e.g.*, Proudfoot, *et al.*, *Science* 209:1329-1336 (1980); U.S. Pat. Nos. 4,683,195 and 4,683,202, and Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory Press, (1989)). Further amplification methods are described *infra*.

Following the extension reaction, extension products are then analyzed to determine whether the extension product includes the labeled nucleotide originally present on the detection primer(s). As described above, the presence or absence of the modified base or label is an indicator of which base is present at the variant site of the target nucleic acid. Initially, extension products are normally separated from unextended primers and other components of the extension reaction. Typically, such separations are achieved using various size-based separation techniques, including, but not limited to, gel electrophoresis and

various chromatographic methods capable of fractionating sample components according to size (*e.g.*, HPLC methods using a sizing column). In some methods, separation and detection are coupled, such as the detection of extension products separated by electrophoresis (*e.g.*, using the Microchannel Plate from Hewlett-Packard). If necessary, unreacted primers can be hydrolyzed using a nuclease such as exonuclease 1. Unreacted primers can also be removed by silica gel membrane filtration, dialysis or using spin columns. In yet other methods, the products can be detected in a homogeneous format without separating product from the excess primers.

Extension products can be analyzed in duplex form or the strands can be separated prior to analysis. If the two strands of the duplex are to be separated prior to analysis, the strands can be separated using appropriate denaturing conditions and methods known in the art including, but not limited to, heat, alkali, formamide, urea, glyoxal, and combinations thereof. In some instances, single-stranded gel separation offers higher resolution, thereby generating a stronger signal, especially in experiments utilizing fluorescence energy-transfer labels (see *infra*). In other instances, the duplex DNA extension product is analyzed directly. This simplifies the method somewhat. Further, as described more fully below, the formation of unlabeled duplex extension product can be detected using various intercalation dyes or by using labeled second primer or detection primers carrying a second label. The ability to detect unlabeled extension product using intercalation dyes can be used to advantage in genotyping experiments, as the presence of labeled and unlabeled extension product indicates that a sample contains two variant forms of the target nucleic acid. Thus, for a diploid organism, this indicates that the organism is a heterozygote for the site being interrogated (see *infra*).

Determination of which nucleotide is present at a variant site depends upon whether the extension product retains the label originally in the detection primer, and, in those instances in which a second primer is utilized to generate amplified product, can also depend on the size of the product. Labeled extension products can be detected directly using methods that are appropriate for the type of label utilized. Thus, for example, if the label is a radioisotope, then a detector capable of measuring radioactivity is utilized; if the label is a fluorophore, then a fluorescence detector is employed. Extension products having a modified base can also be detected in a straightforward manner by labeling the modified base after the extension reaction or by detecting the modified nucleotide directly. For example,

biotinylated nucleotide can be detected by streptavidin coupled to an enzyme or dye. Other aspects and options concerning label selection and the detection process are described further below.

In some methods, the separation of extension product from other components of the reaction mixtures and the actual detection of labeled components is performed on a single integrated device capable of performing both the separation and detection steps. Various suitable instruments capable of performing such analyses are available including, for example, the Micro-Channel Plate available from Hewlett-Packard and MegaBACE from Molecular Dynamics, or ABI Prism Sequencers from PE Biosystems (see also, Woolley, A.T. and Mathies, R.A., *Proc. Natl. Acad. Sci. USA*, 91:11348-11352 (1994); Shi, Y. *et al., Anal. Chem.* 71:5354-5361 (1999); and Quesada, M.A. *et al., BioTechniques*, vol. 10, no. 5, (1991)).

The methods can achieve high signal to noise ratios. The use of proofreading polymerases means that allele discrimination is achieved first because the proofreading capability of the polymerase results in the formation of labeled and unlabeled product depending upon the nucleotide present at the variant site of the target nucleic acid. Discrimination is also achieved because of kinetic competition between the primer that is fully complementary to the nucleotide at the variant site and the primer that is not. The reduction in the rate at which the mismatched primer is extended as compared to the perfectly matched primer further enhances discrimination between alleles. Additionally, as noted above, by labeling an internal primer nucleotide, extension reactions are more efficient because there is less likelihood of the label interfering with the extension reaction. Because the methods are capable of achieving high signal to noise ratios, in some instances it is not necessary to pre-amplify the target nucleic acid prior to conducting an analysis, although the target can be pre-amplified to further enhance the signal to noise ratios.

B. Methods with more than two allelic forms

The foregoing description applies for the most typical situation in which a variant site potentially includes two different nucleotides. In some instances, the variant site can include three or four different nucleotides. The methods of the invention can be easily adapted to such situations. When the sample includes target nucleic acids that include three or four allelic forms, extension reactions are conducted with a detection primer for each of

the polymorphic forms, where each detection primer is capable of being uniquely labeled so that the different allelic forms can be discriminated. As described above, labeled extension product is formed if the nucleotide in the detection primer that aligns with the variant site is complementary to the nucleotide at that site.

C. Multiplexing

The fact that the methods of the invention generate labeled products allows for the facile extension of the basic method described above to multiplexing formats in which the identity of a nucleotide at multiple variant sites is determined in a single reaction. Such formats allow for rapid sequence determinations in many loci and/or individuals simultaneously. Essentially this is achieved by pooling several single nucleotide extension reactions into a single reaction. The multiple sites can be multiple sites on the same target nucleic acid, such sites being within the same gene or at sites in different genes. For example, the multiple sites can be on a single nucleic acid molecule or different sites on the genomic DNA from a particular individual.

The multiplexing methods closely parallel the methods just described. Detection primers for each of the different variant sites are annealed to their respective binding sites. The general structure and labeling of each of the primers and the performance of the extension reactions is as set forth above. In the multiplex format, each primer potentially generates a labeled or unlabeled extension product. As described above, labeled extension product is produced if the detection primer is complementary to the nucleotide at the variant site, whereas unlabeled extension product is formed if the detection primer is not complementary to the nucleotide at the variant site.

In order to correlate the multiple extension products with the various sites, a number of different strategies can be utilized to aid in determining which extension product corresponds with which variant site. One option is to differentially label the primers for the multiple sites, *i.e.*, to use different labels for the different variant sites. Most typically, this means that fluorescent dyes that fluoresce at different wavelengths are utilized for the different detection primers.

Another approach involves using certain primers such that the extension products formed are of differing size and capable of being distinguished by size fractionation. Typically, this is achieved through the judicious selection of target and second primers, such

that different sized extension products are formed for the various sites. Different sized primers can also be used to control the size of the final extension product.

Alternatively, the primers can be tagged with distinctive identifier tags for each variant site to enable the different extension products formed in a multiplexing experiment to be differentiated. A number of different tags can be utilized. In some instances, a secondary label is used in conjunction with the label used to identify the nucleotide present at a particular variant site. Such secondary labels can be any type of molecule or compound that is detectable (see *infra*). In other instances, the tag is part of an affinity pair and facilitates separation of the different extension products one from the other. The components of an affinity pair generally include one agent and a complementary agent capable of specifically binding to one another. Examples of such binding pairs include, but are not limited to, polynucleotide/complementary polynucleotide, biotin/avidin, antigen/antibody and heavy metal/thiol group. In some instances, one member of the affinity pair is attached to a solid support. A solution containing (or potentially containing) the complementary member of the affinity pair is then contacted with the support. After allowing the two components an opportunity to bind and form a complex, other species in the extension reaction mixture, including extension products bearing non-complementary affinity components, can be washed from the column.

A variety of different types of supports can be utilized in those methods employing affinity binding pairs. Suitable supports include, but are not limited to, beads, microparticles, the surface of a microtiter well, a filter and a glass slide. Similarly, the supports can be formed from any material stable to the binding and washing conditions including, for example, glass, polystyrene, cellulose, latex, nitrocellulose, nylon, polyacrylamide, dextran and agarose.

The methods utilizing primers of different size or tagged primers can be used in conjunction with the scheme in which different labels are used for different variant sites. In this way, extension products can be identified and distinguished from one another on the basis of two criteria rather than simply one criterion. For example, in some methods, the detection primers can bear labels and a tag, the detection primers for each variant site including a different label and a different tag. The different extension products can then be identified and/or separated both according to the different tags and the different labels.

D. Pooling studies

The method can be utilized to determine the allele frequency of a variant site in a study population. Typically, in these type of experiments, the DNA samples from different individuals are pooled together. Then the method of this invention can be used to analyze the presence of each allele in the mixed templates. By comparing the signal intensities of each allele with a reference set (for example, the heterozygotes, the homozygotes or a mixture of both at a known ratio), the prevalence of the alleles in the population can be determined. (For a general discussion of pooling studies see, *e.g.*, Breen G. *et al.*, *BioTechniques* 28:464-468 (2000); Risch N. and Teng, J., *Genome Res.* 8:1273-1288 (1998); Shaw, S.H. *et al.*, *Genomer Res.* 8:111-123 (1998); and Scott, D.A. *et al.*, *Am. J. Hum. Genet.* 59:385-391 (1996)).

IV. Samples

A. Types of Target Nucleic Acids

The methods of the present invention can be utilized to determine the identity of a nucleotide at a variety of different types of variant sites including, but not limited to, SNPs and mutations such as transitions, transversions, insertions and deletions. The presence or absence of a target nucleic acid in a sample can be detected generally as the presence or absence of a particular nucleotide. Individual nucleotides located at a particular site can also be identified by the methods described herein.

The methods presented are generally applicable to all DNA sequences, whether the DNA molecules are single- or double-stranded, provided the target nucleic acid is of sufficient length to form a hybrid with a complementary detection primer. RNA samples can be utilized if first reverse transcribed to form cDNA.

The target nucleic acid can be only a fraction of a larger nucleic acid or can be present initially as a purified and discrete molecule. Additionally, the target nucleic acid can constitute the entire nucleic acid or can be a fraction of a complex mixture of nucleic acids. The target nucleic acid can be synthesized enzymatically *in vivo*, synthesized enzymatically *in vitro*, or synthesized non-enzymatically. In some instances, the target nucleic acid can include nucleotide analogs that are not naturally-occurring including, but not limited to, deoxyinosine and 7-deaza-2-deoxyguanosine. Such analogs destabilize DNA duplexes and

can allow primer annealing and extension reactions to occur in double-stranded DNA without completely separating the strands comprising the duplex.

B. Sources

The target nucleic acid can be from any source. The samples that include the target nucleic acids can be natural or synthetic using enzymatic or synthetic organic techniques. Likewise, the sample can be taken from any organism, including but not limited to, plants, microorganisms (*e.g.*, bacteria, fungi and viruses), vertebrates, invertebrates and mammals (*e.g.*, humans, primates, horses, dogs, cows, pigs and sheep).

For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. Samples can be obtained from the tissues or fluids of an organism; samples can also be obtained from cell cultures, tissue homogenates or synthesized as described above. For example, samples can be obtained from whole blood, serum, semen, saliva, tears, urine, fecal material, sweat, buccal, skin, spinal fluid and hair. Samples can also be derived from *in vitro* cell cultures, including the growth medium, recombinant cells and cell components. For assay of cDNA or mRNA reverse transcribed to form cDNA, the tissue sample is obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source. Samples for use in prenatal testing can be obtained from amniotic fluid.

The target nucleic acid(s) can also be obtained from non-living sources suspected of containing matter from living organisms. For example, in the instance of samples obtained for forensic analysis, the target nucleic acids can be obtained from samples of clothing, furniture, weapons and other items found at a crime scene.

C. Sample Preparation

In some instances, the samples contain such a low level of target nucleic acids that it is useful to conduct a pre-amplification reaction to increase the concentration of the target nucleic acids. Pre-amplification of the target nucleic acid can be accomplished, for example, by conducting PCR reactions with primers that flank the detection primer and second/reverse primer binding sites. As noted above, however, certain features of the present methods result in increased signal to noise ratios, thereby frequently making a pre-amplification step unnecessary. If a pre-amplification step is conducted, the linear and

exponential amplification methods described *supra* can be utilized to increase the initial concentration of target nucleic acid. Further guidance regarding nucleic sample preparation is described in Sambrook *et al., Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, (1989)).

V. Extension Reactions

A. Primers

The primers (first/detection primer and second primer, if utilized) can be either naturally occurring nucleic acids or prepared using synthetic methods. If synthesized, the primers can be synthesized either enzymatically *in vitro*, enzymatically *in vivo* or non-enzymatically *in vitro*.

The detection primer and primer for the complementary strand (for exponential amplification reactions) are sufficiently long to specifically bind to the appropriate target nucleic acid and to form a stable hybridization complex under the extension reaction conditions. Typically, the primers are 15 to 50 nucleotides in length; in other instances, the primers are 20 to 30 nucleotides long. The length of the primers can be adjusted to be longer or somewhat shorter depending upon the particular sequence to which they hybridize (*e.g.*, primers with a high G/C content typically can be shorter than those with a low G/C content).

As noted above and depicted in FIG. 1A, the detection primer is designed such that after annealing to the template strand it extends beyond the variant site in the template strand. The length of the segment of the primer that extends 3' of the nucleotide aligned with the variant site in the template strand can vary and is referred to herein as the "3' terminal segment." (see FIG. 1B). The 3' terminal segment is at least 1 nucleotide in length since the 3' terminal nucleotide is not labeled in the present invention. In certain methods, the 3' terminal segment is at least 2 but less than 10 nucleotides in length, in other instances, at least 2 but less than 5 nucleotides in length, and in still other instances, 1 or 2 nucleotides in length.

Most typically, the primers are designed to be perfectly complementary over their entire length with the template strand, with the exception of course that the detection primers may not be complementary with the particular nucleotide at the variant site being interrogated. In some methods, a certain number of additional mismatches can be tolerated

once the primers have hybridized to their primer sites. In general, however, such mismatches in the detection primer should be placed at a position so that it has no effect on the proofreading process.

In certain methods, the detection primer and/or the second primer can include one or more moieties that allow for the affinity separation of the extension product or primer from unincorporated reagents and/or the target nucleic acid and/or other nucleic acids in the test sample. For example, the primer can include biotin which permits the affinity separation of the primer or extension product from other reaction components through binding of biotin to streptavidin molecules attached to a solid support. As another example, a support can be attached to a nucleic acid sequence that is complementary to the primer or extension product generated therefrom. Hybridization between the primer and its complementary sequence also allows for affinity separation.

In some instances, the detection primer includes a pair of fluorescent energy transfer labels, wherein one label serves as a donor dye and the other label serves as an acceptor dye. In such instances, one label is attached to the label attachment region (see FIG. 1B and accompanying text) and the other is attached at a location where it is not susceptible to removal during the proofreading process. Such primers can be used in homogeneous assays in which separation of products is unnecessary. For instance, the detection primer can carry a quencher at the 3' label attachment site (*i.e.*, within the label attachment region), as well as a fluorescent dye 5' to the label attachment site. The quencher and the dye are located within energy transfer distances. During the proofreading amplification process, the quencher is removed if there is a mismatch between the detection primer and the template nucleic acid at the variant site. The removal of the quencher can be detected by the increased fluorescent intensity of the fluorescent dye. Examples of suitable energy transfer dye pairs are set forth *infra* on the section on labels.

B. Amplification

As described above, the methods of the invention frequently involve either a linear or exponential amplification process to enhance signal to noise ratios. If double stranded, the target nucleic acid is first denatured to form single-stranded nucleic acid using any of a variety of denaturation techniques which are known in the art, including, for example, physical, chemical, enzymatic or thermal means. Typically, strand separation is

achieved using heat denaturation at temperatures ranging from 80 °C to about 105 °C for time periods ranging from about 30 seconds to 15 minutes. For cases in which the nucleic acid is RNA, the sample is first reverse transcribed to form cDNA which is then denatured.

The resulting denatured nucleic acid strands are incubated with the primers under hybridization conditions, *i.e.*, conditions in which the primers anneal to their respective complementary portions of the single-stranded nucleic acid. Because the denatured nucleic acid strands are typically considerably longer than the primers, there is an increased probability that a denatured strand makes contact and reanneals with its complementary strand before the primer or probe has a chance to hybridize to their complementary sequences. To avoid this problem, typically a high molar excess of primer is used to increase the likelihood that the primer anneals to its template strand before the denatured strands reanneal.

Amplification is typically conducted using the polymerase chain reaction (PCR) according to known procedures. See generally, *PCR Technology: Principles and Applications for DNA Amplification* (H.A. Erlich, Ed.) Freeman Press, NY, NY (1992); *PCR Protocols: A Guide to Methods and Applications* (Innis, *et al.*, Eds.) Academic Press, San Diego, CA (1990); Mattila *et al.*, *Nucleic Acids Res.* 19: 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1: 17 (1991); *PCR* (McPherson *et al.* Ed.), IRL Press, Oxford); and U.S. Patent Nos. 4,683,202 and 4,683,195. Other suitable amplification methods include: (i) the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4:560 (1989) and Landegren *et al.*, *Science* 241:1077 (1988)); (ii) transcription amplification (see Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86:1173 (1989)); (iii) self-sustained sequence replication (see Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA*, 87:1874 (1990)); and (iv) nucleic acid based sequence amplification (NASBA) (see, Sooknanan, R. and Malek, L., *BioTechnology* 13: 563-65 (1995)).

VI. Detection

A. Multiple Covalently Labeled Detection Primers

As illustrated in FIG. 1D, certain methods involve the use of multiple labeled detection primers in which a detection primer bearing a different label within the label attachment region is used for each allelic form for a variant site. This means that a differentially labeled extension product is formed for each allelic form. Because a labeled

extension product is formed for each allelic form, identification of the different allelic forms present in a sample can rapidly be identified.

An extension product lacking the label originally in the detection primer is also formed for each allelic form (when the detection primer is not complementary to the template at the variant site). Even the unlabeled extension products can be labeled using labeled reverse primers or dual labeled detection primers as described in the following sections.

B. Labeled Second Primer

For methods wherein the target nucleic acid is amplified using a labeled detection primer and a second primer, the second primer can optionally be labeled (see FIG. 5). In such instances, two types of labeled product are formed. When the detection primer is complementary to the target nucleic acid, then the amplification product retains the label originally part of the detection primer and the label borne by the second primer. If the detection primer is not complementary to the target nucleic acid, then the amplification product bears only the label from the second primer, the label from the detection primer having been removed during the proofreading process. Thus, the genotype can be determined on the basis of the ratio of the signals from the two dyes.

As shown in FIG. 5, with this approach, even amplification product in which the label borne by the detection label is excised during the proofreading process is nonetheless easily detected because of the label on the second primer. When the second primer is labeled, essentially any nucleotide can be labeled. For convenience, typically the 5' nucleotide is labeled. A specific example of this approach is described in Example 2 infra.

C. Homogeneous Detection

In certain methods, amplification products are detected in a homogeneous format that does not require labeled amplification product to be separated from labeled nonextended primers and nucleotides prior to determining whether the label in the detection primer (and optionally the second primer) is retained. This enables the assays to be conducted more conveniently and rapidly.

One such approach is illustrated in FIG. 6. In this particular homogeneous assay, the detection ("forward") primer is labeled with one label (e.g., a fluorescent dye).

While the label in FIG. 6 is shown attached to the nucleotide that aligns with the nucleotide at the variant site, the label can be attached at any nucleotide within the label attachment region (see FIG. 1B). The reverse primer can optionally be labeled at essentially any nucleotide. As shown in FIG. 6, it can conveniently be labeled at the 5' end. After amplification with a DNA polymerase with the proofreading activity, the excess labeled primers are degraded with an exonuclease such as Exonuclease I. The whole mixture is then analyzed for fluorescence polarization (FP) on a plate reader.

Fluorescence polarization is a measurement of the rotation rate of a fluorescent molecule. In general, a large molecule rotates slower than a small molecule and hence has a higher FP reading. Thus, FP can be utilized to distinguish between fluorescently labeled molecules of different sizes. In assays such as described herein, since the PCR products are much larger in size than the dye attached to a free nucleotide, the change in fluorescence polarization can provide a measure of the extent to which primers have become incorporated into amplification products. In those instances in which both the detection and reverse primers are labeled, one can infer the genotypes of the samples analyzed on the basis of the relative signals from each label (see Example 2 infra).

This approach can be utilized in various other formats as well. For example, as described above, unlabeled reverse primers can be utilized in combination with differentially labeled detection primers in which each detection primer is complementary to a different allelic form of the target nucleic acid.

D. Multiple Covalent Labels on Single Detection Primer

As described above, typically the detection primer includes a single covalently attached label. However, in a modification of the foregoing methods that utilize multiple labels, in some instances the detection primer includes two or more covalently attached labels. In such instances, the labels are located on the detection primer such that no label is attached to the 3' terminal nucleotide. One of the labels, as described above, is located in the segment of the detection primer that includes the nucleotide aligned with the variant site in the template strand and extends to the nucleotide immediately 5' of the 3' terminal nucleotide (*i.e.*, the label attachment region; see FIG. 1B). It is this label whose presence or absence indicates whether the detection primer is complementary to the nucleotide at the variant site or not. The second label is positioned 5' of the first label and is located at a nucleotide at

which it cannot be removed during the proofreading process. Typically, the second label is located to be within an energy transfer relationship with the first label. Hence, usually, one of the labels is selected to be a donor dye and the other an acceptor dye. Suitable fluorescent dyes can be selected from those set forth below. In some instances, the second label can be further away from the first label. For example, the second label can be attached to the 5' of the detection primer.

E. Covalent Labels in Combination with Intercalation Dyes

In some instances, the extension products are separated and then labeled with non-specific dyes to detect all the extension products. In this way, one can detect not just extension product that bears a covalent label, but those extension products in which the label is excised during the extension reaction because of a mismatch at the variant site. The covalent label refers to the label attached to the primer to distinguish between the nucleotides at a variant site. By conducting these additional steps, one can distinguish between a failed experiment and the situation in which an unlabeled extension product is formed. One approach for conducting this type of analysis is to separate the extension products by size using gel electrophoresis. The products can then be visualized using a variety of dyes that bind to DNA. Essentially any intercalation dye capable of intercalating into duplex DNA can be used to detect extension product. Specific examples of suitable dyes include, but are not limited to, thiazole orange, ethidium bromide, propidium iodide, chromomycin, acridine orange, Hoechst 33258, TOTO-1, YOYO-1, DAPI (4',6-diamidino-2-phenylindole hydrochloride), SyberGreen and Pico Green (the latter two dyes being available from Molecular Probes, Inc. of Eugene, OR).

Certain methods involve a modification of this approach in which unlabeled extension products are labeled with an intercalation dye while all the extension products remain in solution. Hence, this approach is simpler than the approach described in the preceding paragraph in which the extension products are first separated from one another using electrophoresis, for example. As just described, labeling of all extension products allows one to differentiate between the situation in which unlabeled product is formed because the detection primer is not complementary to the nucleotide at the variant site and the situation in which no extension product is formed because of failure of the extension reaction. By staining in this way, one can also observe and distinguish between target nucleic acids

that include different alleles (*i.e.*, between different allelic forms). This ability is important in the genotyping methods described below in which the presence or absence of different allelic forms is determined.

When unlabeled extension products are labeled in solution using intercalation dyes, assays are conducted as described above to determine the identity of the nucleotide at the variant site. However, when the double-stranded DNA sample is separated, the nucleic acids are run in a native buffer containing typically from 10^{-7} to 10^{-9} M intercalation dye. Since the intercalation dye intercalates into DNA with little sequence specificity, extension products generated by both allelic templates are labeled with the intercalation dye and, depending on the concentration of intercalation dye introduced, the number of intercalation dye molecules bound to the DNA can range up to about 1 dye molecule per 5-10 base pairs. Thus, by changing the intercalation dye concentration, the number of intercalation dye molecules bound per DNA extension product can be facilely adjusted so that the emission intensity of the extension product labeled with just intercalation dye is similar to that of the DNA extension product that contains the covalent label.

The fact that the intercalation dye also intercalates into the extension product that bears the covalent dye can be utilized to advantage in distinguishing between the extension products and in enhancing the emission from the covalently labeled extension product, which results in a concomitant increase in the signal to noise ratio. To enable differentiation between the intercalation dye and the covalent label when it too is a fluorescent dye, the covalent dye typically is chosen to emit to the red of the intercalation dye (*i.e.*, the covalent label should emit at a longer wavelength than the intercalation dye). When the dyes are selected in this way, the intercalation dye molecules (donors) that intercalate into an extension product already containing the red-emitting dye (*i.e.*, the covalently attached fluorescent label) can transfer their energy via a Foerster mechanism to the covalently attached fluorescent label (acceptor). As a consequence of this energy transfer, emission from the covalently attached fluorescent label can be enhanced from 5-10 fold or more. Further, the ratio of the emissions from the intercalation dye and the covalently attached fluorescent dye can be utilized to differentiate between extension products that bear the covalent label and those products that do not, even in mixtures of the two products. As described further *infra*, this can be used to advantage in conducting genotyping experiments because the

different emissions allow extension products formed from two alleles to be easily distinguished.

Efficient energy transfer via the Foerster mechanism requires that the donor and acceptor dyes have a certain spacing between the dyes. In general, the transfer of optical excitation from the donor to the acceptor depends upon $1/R^6$, where R is the distance between the two fluorophores. The number of intercalation dye molecules incorporated within a DNA duplex can be regulated by controlling the amount of intercalation dye contacted with the duplex. In general, dyes that are separated from 1 to 20 nucleotides between the covalent dye (dye originally attached to detection primer) and the intercalated dye allow for effective energy transfer; and those donor and acceptor dyes with a separation of between 2 to 10 nucleotides permit the greatest degree of energy transfer. The donor and acceptor dyes should have overlapping spectra, such that the donor emission spectrum overlaps the acceptor absorption spectrum so that there is efficient energy transfer from the excited donor fluorophore to the acceptor fluorophore. Further, in general the emitting or accepting fluorophores are selected to be able to receive the energy from the donor fluorophore(s) and emit light at a distinctive and detectably different wavelength.

A variety of intercalation dyes can be utilized to label the extension products. Suitable intercalation dyes include, but are not limited to, thiazole orange (TO), TOTO, YOYO, TOTAB, EthD, TO6, YO, Eth, propidium-2 and propidium-3. TO is useful because is has a greater than 1000-fold enhancement of its fluorescence upon binding to double-stranded DNA. TO can be excited well at 488 nm and fluoresces near 530 nm. With these absorption and emission characteristics, TO is well-suited for transferring energy to covalently attached fluorescent labels that emit at longer wavelengths than TO. Suitable fluorescent acceptor dyes that can be utilized with TO include R6G, TAMRA, ROX, HEX, JOE and TET.

Additional discussion regarding the use of intercalation dyes is provided by Zhu *et al., Anal. Chem.* 66:1941-1948 (1994).

F. Labels and Modified Nucleotides

1. Types

The detection primer includes either a label or a modified nucleotide that is either directly or indirectly detectable. The label covalently attached to the detection primer

can be any compound or molecule that can be detected and that does not significantly interfere with the extension reaction. Suitable labels include, but are not limited to, fluorophores, chromophores, molecules that chemiluminesce, magnetic particles, radioisotopes, mass markers, electron dense particles, enzymes, cofactors, electrochemically active molecules, substrates for enzymes and ligands having specific binding partners (*e.g.*, avid/biotin). The modified nucleotide typically is modified so that after the primer synthesis or extension reaction the modified nucleotide can be attached to a label such as those just listed. As described more fully below, a modified nucleotide can be a standard nucleotide to which a linker is attached, which in turn is attached to a label. Alternatively, the primer can incorporate a labeled nucleotide directly during synthesis.

Certain methods utilize fluorescent molecules as the labels, as a number of commercial instruments have been developed for the detection of fluorescently labeled nucleic acids. A variety of fluorescent molecules can be used as labels including, for example, fluorescein and fluorescein derivatives, rhodamine and rhodamine derivatives, naphythylamine and naphthylamine derivatives, cyanine and cyanine derivatives, benzamidizoles, ethidiums, propidiums, anthracyclines, mithramycins, acridines, actinomycins, merocyanines, coumarins, pyrenes, chrysenes, stilbenes, anthracenes, naphthalenes, salicyclic acids, benz-2-oxa-1-diazoles (also called benzofurazans), fluorescamines and bodipy dyes.

For those methods in which the detection primer and/or the detection product are labeled with fluorescent dyes capable of energy transfer to enhance emission intensities or simplify the assay, a number of donor (or reporter) and an acceptor (or quencher) dyes are available. One group of donor and acceptor dyes includes the xanthene dyes, such as fluorescein dyes, and rhodamine dyes. A variety of derivatives of these dyes are commercially available. Often functional groups are introduced into the phenyl group of these dyes to serve as a linkage site to an oligonucleotide. Another general group of dyes includes the naphthylamines which have an amino group in the alpha or beta position. Dyes of this general type include 1-dimethylaminonaphthyl-5-sulfonate, 1-anilino-8-naphthalende sulfonate and 2-p-touidinyl-6-naphthalene sulfonate. Other dyes include 3-phenyl-7-isocyanatocoumarin, acridines, such as 9-isothiocyanatoacridine and acridine orange, pyrenes, bensoxadiazoles and stilbenes. Additional dyes include 3-(ε-carboxypentyl)-3'-ethyl-5,5'-dimethyloxa-carbocyanine (CYA); 6-carboxy fluorescein (FAM); 5&6-

carboxyrhodamine-110 (R110); 6-carboxyrhodamine-6G (R6G); N',N',N',N'-tetramethyl -6-carboxyrhodamine (TAMRA); 6-carboxy-X-rhodamine (ROX); 2', 4', 5', 7', - tetrachloro – 4 – 7 – dichlorofluorescein (TET); 2', 7' – dimethoxy – 4', 5' – 6 carboxyrhodamine (JOE); 6-carboxy-2'4,4',5',7,7'-hexachlorofluorescein (HEX); ALEXA; Cy3 and Cy5. These dyes are commercially available from various suppliers such as Applied Biosystems Division of Perkin Elmer Corporation (Foster City, CA), Amersham Pharmacia Biotech (Piscataway, NJ), and Molecular Probes, Inc. (Eugene, OR).

Further guidance regarding the selection of donor and acceptor pairs that can effectively be used with the methods of the present invention include: *Fluorescence Spectroscopy* (Pesce *et al.*, Eds.) Marcel Dekker, New York, (1971); White *et al.*, *Fluorescence Analysis: A Practical Approach*, Marcel Dekker, New York, (1970); Berlman, *Handbook of Fluorescence Spectra of Aromatic Molecules*, 2nd ed., Academic Press, New York, (1971); Griffiths, *Colour and Constitution of Organic Molecules*, Academic Press, New York, (1976); *Indicators* (Bishop, Ed.). Pergamon Press, Oxford, 19723; and Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, Molecular Probes, Eugene (1992).

2. Label Attachment

Attaching a label to a modified nucleotide can be achieved in a number of different ways. One general approach involves preparing derivatives of dyes that contain appropriate functional groups for linking the dyes to the detection primer. Such methods are described, for example, by Marshall, *Histochemical J.* 7:299-303 (1975); Mechnen *et al.* in U.S. Pat. No. 5,188,934; Bergot *et al.* in PCT publication PCT/US90/05565; Ullman *et al.* in U.S. Pat. No. 3,996,345 and Khanna *et al.* in U.S. Pat. No. 4,351,760.

In another approach, a label is linked to a nucleotide in the detection primer via a linker. A number of such linkers are commercially available and have varying lengths. Such linkers are useful for obtaining a desired distance between the primer and label to ensure that the label does not interfere with the extension reactions. In general such linkers include a functional group (*e.g.*, amino, hydroxyl, sulfhydryl, carboxyl) at each end so that one end can be attached to a nucleotide in the detection primer and the other end attached to the label (*e.g.*, fluorescent molecule). Examples of such linkers include "Amino Modifier C3", "Amino Modifier C6," "Amino Modified C7" and "Amino Modified C12" that are

available from Operon Technologies, Inc. (see FIG. 5). Another suitable linker is the "Uni-Link Amino Modifier" available from Clonetech (Palo Alto, CA) (see FIG. 5).

Alternatively, modified nucleotides designed to allow for attachment of a label can be incorporated into the detection primer during synthesis. Examples of such modified nucleotides include, for example, 5'-dimethoxytrityl-5-[N-(trifluoroacetylaminohexyl)-3-acrylimido]-2'-deoxyuridine,3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite, marketed under the name "Amino-Modifier C6 dT" and similar modified nucleotides available from Glen Research (Sterling, VA) and designed to function as modified thymidine nucleotides (see, *e.g.*, "Users Guide to DNA Modification," Glen Research Corporation, pp. 1-68 (1996)). Similarly modified deoxycytidine compounds can also be utilized (see, *e.g.*, Markiewicz,W.T., *et al., Nucleosides and Nucleotides*, 11:1703-1711 (1992)). These molecules contain a protected primary amine that can serve as the attachment site of a label (*e.g.*, a fluorescent label) following deprotection. Methods for incorporating such modified nucleotides into a primer are described, for example, in U.S. Pat. Nos. 5,654,419; 5,688,648; 5,853,992; and 5,728,528 to Mathies *et al.* and Glazer *et al.* Single dye labeled R110-ddNTPs, R6G-ddNTPs, TAMRA-ddNTPs and ROX-ddNTPs can be purchased from DuPont NEN (Boston, MA).

In other instances in which label is attached during synthesis, prelabeled nucleotides (*e.g.*, TAMRA-dT or Dabcyl-dT can be obtained commercially (e.g., Glen Research, Sterling, Virginia)) are obtained in functionalized form (*e.g.*, as a phosphoramidite) and are incorporated during organic synthesis. Radioactive label can be incorporated in a similar manner.

VII. Genotyping

A diploid organism contains two copies of each gene. Genotyping involves the determination of whether a diploid organism contains two copies of the reference allele (a reference-type homozygote), one copy each of the reference and variant allele (*i.e.*, a heterozygote), or contains two copies of the variant allele (*i.e.*, a variant-type homozygote). When conducting a genotyping analysis, the methods of the invention can be utilized to interrogate a single variant site. However, as described further below in the section on

multiplexing, the methods can also be used to determine the genotype of an individual in many different DNA loci, either on the same gene, different genes or combinations thereof.

Most typically, SNPs consist of two allelic forms, *i.e.*, the variant site includes one of two different nucleotides. This is the situation illustrated in FIG. 1C. Referring again to this figure, if a sample from a diploid individual is obtained, the sample can contain nucleic acids representative of the two copies of the target nucleic acid of interest. If the sample is obtained from a homozygote, either reference or variant type, and a single detection primer is used to conduct the extension reactions, a single extension product is formed. More specifically, if the detection primer is complementary to the nucleotide at the variant site, then only extension product retaining the label in the detection primer is formed. Utilization of a detection primer lacking complementarity to the nucleotide at the variant site, in contrast, yields an extension product in which the label from the detection primer is removed during proofreading. Thus, regardless of the specific product formed, for a homozygote, only a single extension product is formed from a single detection primer.

If, however, the sample is from a heterozygote, then the sample includes target nucleic acids with two different nucleotides at the variant site. For an analysis conducted with a single detection primer, then both types of primer/template hybridization complexes shown in FIG. 1C are formed, namely a complex in which the detection primer is complementary to the nucleotide in the variant site of the template and a complex wherein the detection primer is not complementary to the nucleotide in the variant site of the other template (*i.e.*, the other allelic form). Hence, an extension product retaining the detection primer label is formed, as well as an extension product lacking the detection primer label. Consequently, whereas a single extension product is formed with a homozygote (whether reference or variant type), two extension products are formed with a heterozygote.

As just described, only a single primer is necessary to determine the genotype of an individual when there are only two alleles (*i.e.*, polymorphic forms). However, using two labeled detection primers, both alleles can be detected directly as shown in FIG. 1D. In this approach, two detection primers, each complementary to one allele, are differentially labeled. After extension, four extension products are formed: two labeled and two unlabeled. The particular labeled detection product(s) formed indicate the presence of the corresponding allele. Similarly, the absence of labeled detection product indicates the absence of the corresponding allele.

In those instances in which only a single detection primer is utilized, unlabeled extension product can nonetheless be detected using the methods described above in which DNA dyes such as intercalation dyes are used to detect the formation of unlabeled extension product. By selecting the covalent label and the intercalation label to have overlapping but distinct spectra and determining a ratio between the emissions from the two labels, one can distinguish extension products with or without the covalent label, even in mixtures of such products.

Alternatively, one can use labeled reverse primer in conjunction with a single labeled detection primer to genotype both alleles. A dual labeled detection primer can be used for this purpose as well. The ratio of signals from the two labels can be used to determine the relative amounts of extension product formed from the two alleles and thus whether the sample is from a homozygote or heterozygote.

The discussion in the preceding paragraph is for the situation in which there are two polymorphic forms. Situations in which there are additional polymorphic forms (*e.g.*, a triallelic case) can be analyzed according to the methods described above in which extension reactions are conducted using a different labeled detection primer for each polymorphic form.

The ability to use the methods of the invention to make rapid genotyping determinations provides a powerful tool in genetic analysis and ascertaining the susceptibility of an individual to a disease. Individuals that are homozygotes for an allele associated with a particular disease are at higher risk of obtaining the disease than a heterozygote or a homozygote for the non-disease associated allele. The heterozygote, however, is a carrier of the allele associated with the disease. Such knowledge can be useful in prenatal and other types of medical and genetic counseling, for example.

VIII. Kits

Kits for conducting the sequence and genotyping determinations described herein are also provided by the invention. Typically, the kits contain one or more detection primers that span a segment that includes the variable site in the template strand of a target nucleic acid. The detection primers can be labeled or simply contain a modified nucleotide that easily permits attachment of a label. If the label is not attached, the kit can also include a quantity of one or more labels for attachment to the detection primers. Other detection

primers include two covalently attached labels, typically energy transfer dyes. Certain kits also include primers (*i.e.*, reverse primers) to be paired with the detection primer for use in amplifying a segment from a target nucleic acid that includes the variant site. The detection primers can be labeled or unlabeled. Such primers are complementary to a downstream segment of the strand opposite that of the template strand. Typically, the primers are designed for use in detecting one or more of the SNPs described in the following section. The number of primers (or primer pairs) included in the kit can vary. Generally, the kits include at least 2, 3, 4 or 5 primers (or primer pairs). Other kits can include more primers, such as at least 10, 15, 20 or 25 primers (or primer pairs).

The kits can include various other components for conducting template-dependent extension reactions including, for example, a polymerase having a 3'-5' exonuclease activity capable of excising mismatches, deoxynucleotide triphosphates, and buffers. Certain kits also include an intercalation dye for use in detecting certain extension products. Kits can also include the necessary electrophoretic components to size separate the extension products formed during an analysis. Such components include gel polymers (*e.g.*, agarose), polymerizing agents and buffers. Typically, the kits also include containers for housing the various components and instructions for using the kit components to conduct an analysis.

IX. Utility

The methods and kits of the invention are generally useful for determining the identity of a nucleotide at a variant site. These methods, however, find use in a variety of more specific applications. One use is the identification and detection of point mutations (*e.g.*, somatic point mutations) or other mutations that result in sequence change, specifically those mutations known to be correlated with diseases. For example, the methods described herein are useful for identifying whether a nucleic acid from a particular subject includes a reference allele or a variant allele at a particular site. Furthermore, in a single analysis, the methods can be utilized to establish the genotype of the individual being tested (*i.e.*, distinguish whether the individual is a homozygote for one or the other alleles, or a heterozygote at a particular polymorphic site (*e.g.*, a SNP site).

The genotyping utility of the present methods makes them useful within the context of medical diagnosis and prognosis. Since many SNPs are associated with various

diseases, clinicians can utilize the results of the genotype study to assess the presence of disease, whether an individual is a carrier of disease, the likelihood that an individual will get a particular disease and the efficacy of various treatment alternatives.

The methods also have a variety of non-medical uses. Such utilities include detecting pathogenic microorganisms, paternity testing and forensic analysis. The methods can also be used to identify SNPs in non-humans, including for example plants, bacteria and viruses. These various uses are described more fully below.

A. Correlation Studies

Use of the methods of the present invention to acquire diagnostic information involves obtaining a sample from a number of different individuals known to have a common disease and conducting screening tests to determine whether they consistently share a common genotype at one or more SNP sites. The results of such screening can be used to establish correlations between certain genotypes and certain diseases.

In a related fashion, the methods of the invention can be used to develop correlations between certain genotypes and patient prognosis. For example, the genotype of a population of individuals suffering from a common disease can be determined at one or more SNP sites. The health history of the individuals can be monitored with time to establish correlations between certain genotypes and disease outcomes. Alternatively, the allele frequencies can be determined in pooled samples using the present invention.

The methods of the invention can also be used to formulate optimal treatment protocols for a particular disease. The methods described herein can be used to place individuals into groups that share a common phenotype and genotype. The group can then be subdivided into various groups that each receive various forms of treatment. By monitoring the health status of the different treatment groups over time, the most effective treatment program for a particular genotype can be established.

B. Use of Current Methods as Screening and Therapeutic Tool

In instances in which a correlation between a particular genotype and disease state have already been established, the methods of the invention can be utilized as a diagnostic tool, a prognostic tool and as a means for assessing the success of various treatment options.

For patients having symptoms of a disease, the methods of the present invention can be used to determine if the patient has a genotype known to be associated with a disease that commonly causes the symptoms the patient exhibits. For example, if the genotyping methods of the invention show that the individual has a genotype associated with a particular disease and further that the genotype is associated with poor recovery (*e.g.*, a homozygote for a variant allele), the physician can counsel the client regarding the likely effectiveness of aggressive treatment options and the option of simply foregoing such treatments, especially if the disease is quite advanced. On the other hand, if the genotype is associated with good recovery, the physician can describe a range of treatment options varying from simply monitoring the disease to see if the condition worsens or more aggressive measures to ensure that the disease is attacked before it gets worse.

The methods of the present invention are also valuable for assessing the actual risk of an individual known to be susceptible to acquiring a disease (e.g., an individual coming from a family that has a history of suffering from a disease). By determining whether the individual is a homozygote for the allele associated with the disease or a heterozygote, a physician can more accurately assess and counsel the patient regarding the likelihood that the patient will begin suffering from disease, factors involved in triggering the disease and the pros and cons regarding different treatment alternatives.

Similarly, certain methods of the invention can also be used to identify individuals at risk for disease, even though they have no symptoms of disease or no known susceptibilities to disease. An individual in this category would generally have no disease symptoms and have no family history of disease. In such cases, the methods of the present invention can be used as a useful preventive screening tool. Using the methods of the present invention, a number of selected SNP sites known to be associated with certain diseases can be interrogated to identify the genotype of the individual at those sites. If a particular genotype were identified that was known to be associated with a particular disease, then a physician could advise the individual regarding the likelihood that the disease would manifest itself and the range of treatment options available.

C. Examples of Diseases that can be Monitored

A large number of diseases have been shown to be correlated with particular allelic forms of SNPs. A large number of such SNPs are listed in WO 93/02216 and by

Cooper et al. (*Hum. Genet.* 85:55-74 (1990)). Specific examples of diseases associated with SNPs include: sickle cell anemia and β-thalassemias (mutation in β-globin gene; Antonarakis, *New Eng. J. Med.*, 320:153-163 (1989)), cystic fibrosis (mutation in cystic fibrosis transmembrane receptor (CFTR); see Kerem, *et al., Science* 245:1073-1080 (1989)), hyperlipoproteinemia (mutation in apolipoprotein E gene; see Mahley, *Science* 240:622-630 (1988)), a wide variety of autoimmune diseases (mutations in human major histocompatibility complex; see Thomson, *Ann. Rev. Genet.*, 22:31-50 (1988); Morel *et al., Proc. Nat. Acad. Sci. USA*, 85:8111-8115 (1988); and Scharf, *et al., Proc. Nat. Acad. Sci. USA*, 85:3504-3508 (1988)) and the formation of oncogenes (mutations to the human ras-gene family; see, e.g., Bos *et al., Nature*, 315:726-730 (1985); Farr *et al., Proc. Natl. Acad. Sci. USA*, 85:1629-1633 (1988); and Neri, *et al., Proc. Natl. Acad. Sci. USA*, 85:9268-9272 (1988)). Other genes containing SNPs associated with disease include genes encoding for angiotensinogen, angiotensin converting enzyme, cholesterol ester transfer protein, dopamine receptors, serotonin receptors, and HIV reverse transcriptase (RT).

D. Other Uses

The methods described herein can also be used to identify point mutations in microorganisms that could potentially cause altered pathogenicity or resistance to certain therapeutics. The methods can also be used to identify cells and strains having a desired genetic constitution for use in various biotechnology applications.

The methods described herein can also detect the presence of somatic mutations that can result in various diseases, including cancer for example.

With knowledge gained from the genotyping capabilities of the methods described herein, clinicians can conduct prenatal testing using cells obtained from a fetus to check for a variety of inheritable diseases, such as those diseases associated with the SNPs listed above. The methods can also be used to identify carriers of mutant alleles. Such information can be of use by a couple prior to conception as they evaluate the risks of having a child with certain birth defects or inheritable diseases.

Methods of the invention may also be utilized in various identification applications, such as in the field of forensic medicine or paternal testing. In the case of forensic analysis, polymorphisms in specific genes can be determined in, for example, blood or semen obtained from a crime scene to indicate whether a particular suspect was involved

in the crime. In like manner, polymorphism analysis may be utilized in paternity disputes to aid in determining whether a particular individual is the parent of a certain child.

In another application, certain methods of the invention are used in blood typing or tissue classification. Tissue classifications, for example, can be determined by identifying polymorphisms specific for a particular individual.

The method can also be used to study gene expression in various cellular states and/or in response to certain environmental factors, for example. Such studies can be conducted after the mRNA is converted to cDNA.

The following examples are provided to illustrate certain aspects of the invention, and should not be construed in any way to limit the scope of the invention.

EXAMPLE 1

Genotyping LPL Polymorphisms

I. Materials and Methods

A. Primer design

The primers were generally chosen to have a melting temperature of approximately 60 °C. The forward primers (detection primers) contained an amino modified pyrimidine near the 3' end for post synthesis labeling. The modified nucleotide was introduced using standard synthesis chemistry and an amino-modified deoxypyrimidine phosphoramidite (see, *e.g.*, "User Guide to DNA Modification," Glen Research, Sterling, Virginia (1996). In the initial experiments, the modified nucleotide corresponded exactly to the sequence variant to be interrogated.

B. Oligonucleotide synthesis and labeling

Oligonucleotides were prepared on ABI 3948 Nucleic Acid Synthesizer (Applied Biosystems) according to the manufacturer's instructions. They were then desalted by cartridge and/or purified by reverse phase HPLC (see, Hung, S-C. *et al., Anal. Biochem.* 255:32-38 (1998)).

For post-synthesis labeling of the polynucleotides containing the amino-modified nucleotide(s), the polynucleotides were first deprotected in ammonium hydroxide

using established procedures (see, *e.g.*, "User Guide to DNA Modification," Glen Research, Sterling, Virginia (1996)). After drying down, the primers were labeled with various fluorescent dye (*e.g.*, TAMRA, ROX or FAM) succinimidyl esters (Molecular Probes) and purified by HPLC (see, Hung, S-C. *et al., Anal. Biochem.* 255:32-38 (1998)).

C. Polymerase-chain-reaction (PCR) with a proofreading polymerase

PCR mixtures containing Pfu polymerase (Stratagene) were set up following the manufacturer's recommendations. Cycling was carried out in a GeneAmp PCR System 9700 (Perkin Elmer) using the following parameters: 94 °C, 2 min. followed by 40 cycles of 94 °C for 15 sec., 55 °C for 1 min., 72 °C for 15-20 sec. As the final step, the reactions were incubated at 72 °C for 5 min. and then left at 4 °C until further processing.

D. Analysis of the PCR products

First, a small fraction of each PCR mixture was analyzed on 3% agarose gel containing Ethidium Bromide in TBE buffer. The results were visualized and recorded under UV light.

For analysis by denaturing gel electrophoresis on a micro-channel-plate (MCP), the PCR products were first purified through a Qiagen PCR purification kit. The DNA was eluted in 2-3 volumes of the elution buffer. After spiking with a labeled oligonucleotide, a 1 μl sample from each was loaded onto the MCP and run on the scanner (see, Shi, Y. *et al., Anal. Chem.* 71:5354-5361 (1999)). The results, recorded as fluorescence intensity versus time, were graphed using Microsoft Excel.

II. Results

A. Genotyping LPL (Lipoprotein Lipase Precursor) SNP-19 T/G polymorphism at position 8393

Ten ng of genomic DNA from five individuals (PD1, 2, 3, 5, and 8) was used as the template (*i.e.*, the target nucleic acid) in separate PCR reactions. The partial nucleotide sequence of LPL surrounding the LPL-19 SNP site is shown in FIG. 2A. The genotype of the individuals was as follows: PD1 (a T homozygote), PD2 (a heterozygote), PD3 (a T homozygote), PD5 (a heterozygote) and PD8 (a G homozygote). The forward primer

(detection primer) contained a FAM-labeled dT residue corresponding to one allelic form of the SNP (*i.e.*, the T allelic form). The label was attached immediately 5' to the 3' end of the primer (see FIG. 2B). The forward and reverse primer binding sites are shown in FIG. 2A.

FIG. 2C shows that a specific PCR product of the expected size was generated in all cases. Results from the MCP scanner analysis (FIG. 2D) indicate that the dye labeled nucleotide was completely removed from resulting PCR products when there was a mismatch between the primer which had a T at the polymorphic site and target nucleic acids from G homozygotes (*i.e.*, individual PD8). Individuals that were T homozygotes or heterozygotes, in contrast, generated labeled extension product. This demonstrated the specificity of the proofreading process.

B. Complete Removal of Mismatched Nucleotide at LPL SNP-9 (A/C polymorphism at position 5554)

The partial sequence of LPL encompassing the LPL-9 SNP site is depicted in FIG. 3A. Templates for the PCR reactions were either genomic DNA from 3 individuals (PD1, 3, and 8) or a positive control sequence that had the same sequence as the other targets, except at the polymorphic site at which a T was introduced by PCR-directed mutagenesis. The genotype of the three individuals at position 5554 was as follows: PD1 (A homozygote), PD3 (A/C heterozygote) and PD8 (C homozygote). The forward primer contained a FAM-labeled dT residue aligned with the SNP site, the SNP site being immediately adjacent the 3' end of the forward primer (see FIG. 3B). A fraction of each reaction was separated by agarose electrophoresis to demonstrate a successful amplification (FIG. 3C).

FIG. 3D shows the results of the analysis of the PCR products as separated on the MCP scanner. Since the polymorphic site of this SNP is occupied by A or C, label should not be retained for either allelic form. This in fact was the observed result. Label was retained in the PCR fragment only when there was a perfect match between the template and the primer (*i.e.*, the "T" allelic form in the positive control). Since none of the target nucleic acids from the test individuals contained T at the polymorphic site (but instead contained A and/or C), no labeled extension product was formed.

C. Proofreading of mismatches that are placed a few nucleotides away from the 3' end

This experiment determined the genotype of PD3 (an A/C heterozygote) for the polymorphism LPL-SNP 9 described in the preceding experiment (see FIG. 3A). Experiments were conducted using primers labeled at different locations. In particular, the forward primer included a Rox-labeled dT at the polymorphic site that was located either one or two nucleotides upstream of the 3' end of the forward primer (see FIG. 4A). Extension reactions were performed with the control sequences described in section (B) of this section wherein a T nucleotide was introduced into the target nucleic acid by PCR-directed mutagenesis and sample from individual PD3. Both primers worked equally well in PCR amplification (FIG. 4B). Further, the proofreading activity of Pfu polymerase successfully removed the labeled dT in the presence of a mismatch (*i.e.*, when the template was from individual PD3 and included the A or C allelic form). The positive controls generated labeled product because they included T at the polymorphic site.

EXAMPLE 2

Two-Color Proofreading Assay using End-Labeled Second Primers with Fluorescent Polarization Detection

I. Background

This set of experiments was conducted in the homogeneous format generally illustrated in FIG. 5 to determine the genotype of individuals at the LPL SNP-5 polymorphic site (an A/T polymorphic site). The partial nucleotide sequence of the LPL surrounding this SNP site is shown in FIG. 7A, with primer binding sites being underlined and the polymorphic site being indicated by the letter "W" (to reflect the fact that the nucleotide can be either A/T). The genotype of the individuals tested was PDRO2 (A/A homozygote), PDR05 (A/A homozygote), PDR03 (A/T heterozygote), PDR06 (A/T heterozygote), PDR08 (T/T homozygote) and HD1.10 (T/T homozygote).

II. Materials and Methods

A. Primer Preparation

Oligonucleotides utilized in primer preparation were purchased from Operon in deprotected form. They were labeled with fluorescent dye succinimidyl esters (Molecular Probes) and purified through reverse phase HPLC (see supra, Example 1).

The forward/detection primer (see FIGS. 7B and 7C) was selected to be perfectly complementary to the sequence variant to be interrogated and contained an amino modified pyrimidine (C6-amino dT) immediately adjacent to the 3' end for post synthesis labeling with the fluorescent dye R110 (blue color) (see FIG. 7B). For the reverse or secondary primers, a C6 amino modifier was added to the 5' end for post synthesis labeling with the fluorescent dye ROX (red) (see FIG. 7C). The binding sites for these two primers are shown in FIG. 7A.

B. Proofreading PCR

PCR mixtures containing Turbo Pfu polymerase (Stratagene) were set up following the manufacturer's recommendations. For negative controls, genomic DNA was replaced by water. Cycling was carried out in a GeneAmp PCR System 9700 (Perkin Elmer) using the following parameters: 94 °C, 2 min followed by 40 cycles of 94 °C for 15 sec., 55 °C for 1 min., 72 °C for 15-20 sec. As the final step, the reactions were incubated at 72 °C for 5 min. and then left at 4 °C until further processing.

C. Removal of Excess Primers by Exonuclease I Digest

To each 10-µl PCR reaction mix, 1 µl (5 units) of Exonuclease I (USB) was added. Incubation was carried out at 37 °C for 30 minutes to 1 hour to allow complete digestion of excess primers. The samples were then kept at 4 °C (for a few hours) or –20 °C (for a few days) until analysis.

D. Analysis of the PCR Products

For detection by fluorescent polarization, each sample was diluted 10x with 10 mM TRIS.Cl, pH 8.5. After transfer to a black 96-well plate, fluorescent polarization (FP) was determined in the Analyst AD plate reader from LJL. Instrument settings were as follows: for R110, excitation=490 nm and emission=520 nm; for ROX, excitation=580 nm and emission=610 nm; a 50/50 beamsplitter was used for both dyes. Each plate was read 3x using same settings to minimize instrument variation.

The raw FP data were analyzed as follows: first, the average of the 3 readings for each sample was determined and these were used for subsequent calculations. Secondly,

the net FP (Δ FP) increases were calculated by subtracting the reading from the negative (i.e., no template) controls. Finally, the ratio of the ΔFP for the 2 dyes was calculated for each PCR sample and plotted to detect the clustering of the three genotypes.

For analysis by electrophoresis, the samples were first desalted using Qiagen PCR Purification kit before running on non-denaturing gel on the MegaBACE (Amersham) (see also Example 1). The data shown are not corrected for spectrum overlap.

The genotypes of the samples analyzed have been confirmed by sequencing.

III. Results

Electropherograms obtained for six different individuals is shown in FIG. 8, with the genotype of the individual at the LPL SNP-5 polymorphic site shown in parentheses. Samples from individuals having a T polymorphic form should produce amplification product containing R110 because amplification reactions were conducted using a detection primer including T at the variant site. Samples from these individuals should also generate ROX labeled amplification product due to ROX labeled second primer. In contrast, amplification reactions conducted with samples from individuals having an A polymorphic form should not yield R110 labeled amplification product. In this instance, label is excised via the proofreading function of the polymerase because the detection primer is not complementary to the nucleotide at the polymorphic site. However, ROX labeled amplification product should still be generated with this polymorphic form. Thus, while R110 labeled amplification product is produced only for T alleles, ROX labeled product is formed for both T and A alleles.

Consequently, A/A homozygotes should generate only a red signal from the ROX label since both copies of the gene produce amplification product in which ROX is retained and R110 excised. With T/T homozygotes, in contrast, both copies of the gene yield amplification product in which both R110 and ROX are retained. Hence, a red signal (ROX) and a blue signal (R110) of approximately the same intensity are generated, assuming approximately similar signal strengths for each label. Finally, with A/T heterozygotes, because both copies of the gene yield ROX-labeled amplification product, whereas only one copy of the gene produces R110-labeled product, the R110 signal (blue) is approximately one-half that of the ROX signal (red). With reference once again to FIG. 8, it can be seen

that these are the results obtained for the six individuals tested, a group that included both types of homozygotes as well as heterozygotes.

A more extensive set of experiments was conducted with 57 different individuals of varying genotypes. When a measure of the ratio of fluorescence polarization (Δ FP-R110/Δ FP-ROX; see supra) for amplified product bearing the two dyes is plotted as shown in FIG. 9, the signals fall into three groups as expected depending upon the genotype of the individual. Of the 57 individuals tested, the genotype of only one individual was non-callable.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent or patent application were specifically and individually indicated to be so incorporated by reference.

WHAT IS CLAIMED IS:

1. A method for analyzing a variant site in a target nucleic acid, comprising:

(a) conducting a sequence-specific amplification step, the amplification step comprising amplifying the target nucleic acid by extending a first and second primer in the presence of a polymerase having a 3'-5' exonuclease activity, the first primer annealing to a segment that spans the variable site in a first strand of the target nucleic acid and bearing a label on at least one nucleotide other than the 3' terminal nucleotide, the second primer being complementary to a segment of a second strand of the target nucleic acid which is complementary to the first strand, whereby

if the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, the first primer is extended to produce a product retaining the label and

if the first primer lacks complementarity to the base occupying the variable site in the first strand of the target nucleic acid, the 3'-5' exonuclease activity of the polymerase digests bases from the 3'-end of the first primer, thereby removing the label from the first primer, and wherein the first primer is thereafter extended to produce a product lacking the label; and

(b) analyzing the product, the analyzing step comprising determining whether the product retains the label to identify the nucleotide at the variant site.

2. The method of claim 1, wherein the amplification step comprises amplification by polymerase chain reaction (PCR).

3. The method according to claim 1, wherein the labeled nucleotide is aligned with the variable site.

4. The method according to claim 1, wherein the labeled nucleotide is 3' to the nucleotide aligned with the variable site.

5. The method according to claim 3, wherein the nucleotide positions in the first primer are numbered consecutively from the 3' end with the 3' terminal nucleotide being designated the first position base, and wherein the labeled nucleotide is at the second, third, fourth or fifth base position.

6. The method according to claim 4, wherein the nucleotide positions in the first primer are numbered consecutively from the 3' end with the 3' terminal nucleotide being designated the first position base, and wherein the labeled nucleotide is at the third, fourth, fifth or sixth base position.

7. The method according to claim 1, wherein the label is selected from the group consisting of a fluorophore, a chromophore, a radioisotope, an enzyme substrate, an electron dense agent, a magnetic particle, a mass label and an electrochemically active molecule.

8. The method according to claim 7 , wherein the label is a fluorescent dye.

9. The method according to claim 8, wherein the fluorescent dye is selected from the group consisting of FAM, ROX, TAMRA, R110, R6G, Alexa, HEX, TET, Joe, Cy3 and Cy 5.

10. The method according to claim 9, wherein the fluorescent dye is FAM, TAMRA or ROX.

11. The method according to claim 1, wherein the polymerase is Pfu, Tli, Vent, Deep Vent or Pfx.

12. The method according to claim 11, wherein the polymerase is Pfu or Pfx.

13. The method according to claim 1, wherein the analyzing step further comprises separating the product from the labeled first primer.

14. The method according to claim 1, wherein the product is a DNA duplex and the analyzing step further comprises separating the product into single stranded DNA and determining whether label is retained in one of the single strands.

15. The method according to claim 1, wherein the analyzing step further comprises separating the product from the primer by a size based separation technique.

16. The method according to claim 15, wherein the separation technique is HPLC or electrophoresis.

17. The method according to claim 16, wherein the separation technique is electrophoresis.

18. The method according to claim 1, wherein

(i) the second primer is labeled with a second label that is distinguishable from the label borne by the first primer; whereby

if the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, then product formed retains both labels and

if the first primer lacks complementarity to the base occupying the variant site in the first stand of the target nucleic acid, then product formed bears only the second label; and

(ii) the analyzing step further comprises determining whether the product bears the first primer label and/or second label.

19. The method according to claim 18, wherein the analyzing step further comprises comparing the signal intensities arising from the first primer label and second label.

20. The method according to claim 18, wherein the analyzing step further comprises cleaving non-extended first and second primers with an exonuclease and detecting product by fluorescence polarization.

21. The method according to claim 1, wherein

(i) the first primer bears a second label at a location other than the nucleotide to which the label is attached, and wherein

if the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, then product formed retains both labels;

if the first primer lacks complementarity to the base occupying the variant site on the first strand of the target nucleic acid, only the second label is retained in the product; and

(ii) the analyzing step further comprises detecting product that retains the label and/or second label.

22. The method according to claim 8, wherein the product is duplex DNA and the analyzing step further comprises:

exposing the product to an intercalation dye to introduce a plurality of intercalation dye molecules into the labeled and/or unlabeled product; and

detecting fluorescence from the fluorescent dye and the intercalation dye.

23. The method according to claim 22, wherein the fluorescent dye is sufficiently close to the intercalation dye in the labeled product such that energy transfer between the fluorescent dye and the intercalation dye can occur, and wherein the analyzing step further comprises:

exciting the intercalation dye, whereby fluorescent emission from the intercalation dye is transferred to the fluorescent dye; and

detecting fluorescence from the fluorescent dye.

24. The method according to claim 23, wherein the intercalation dye is thiazole orange.

25. The method according to claim 23, wherein the label is a fluorescent dye that fluoresces at a longer wavelength than the intercalation dye.

26. The method according to claim 1, wherein the target nucleic acid is from a diploid subject and potentially comprises a plurality of nucleic acids in which the nucleotide at the variant site differs, whereby if the sample contains the plurality of nucleic acids the first and second primer generate labeled and unlabeled product, and wherein

the generation of only labeled product or only unlabeled product indicates that the subject is a homozygote, and

the generation of labeled and unlabeled product indicates that the subject is a heterozygote.

27. The method according to claim 1, wherein the target nucleic acid is from a diploid subject and potentially comprises a plurality of nucleic acids in which the nucleotide at the variant site differs, whereby if the sample contains the plurality of nucleic acids the first and second primer generate labeled and unlabeled product, and wherein the label is a fluorescent label and the analyzing step further comprises:

(i) exposing the product(s) to an intercalation dye to introduce a plurality of intercalation dye molecules into the labeled and/or unlabeled product;

(ii) detecting fluorescence from the intercalation dye and fluorescent label, wherein

the detection of only product bearing the fluorescent label and the intercalation dye or only product bearing just intercalation dye indicates that the subject is a homozygote, and

the detection of product bearing both the fluorescent label and the intercalation dye and product bearing just intercalation dye indicates that the subject is a heterozygote.

28. The method according to claim 27, wherein the analyzing step further comprises determining the ratio between emission intensity from the fluorescent dye and the intercalation dye.

29. The method according to claim 1, further comprises cleaving non-extended first primer with an exonuclease and detecting product by fluorescence polarization.

30. A method for analyzing a variant site in a target nucleic acid, comprising:

(a) conducting a sequence-specific amplification step, the amplification step comprising amplifying the target nucleic acid in a sample containing multiple copies of the target nucleic acid by extending a plurality of primer pairs in the presence of a

polymerase having a 3'-5' exonuclease activity, each primer pair comprising a first and second primer, wherein

(i) each of the first primers anneals to a segment that spans the variable site in a first strand of the target nucleic acid and bears a label on at least one nucleotide other than the 3' terminal nucleotide, and the first primers being selected to be complementary to each of the allelic forms of the target nucleic acid such that the first primers in the different primer pairs are complementary to a different one of the plurality of bases that potentially occupy the variant site and bear different labels;

(ii) the second primer of each pair is complementary to a segment of a second strand of the target nucleic acid; and

(iii) the first and second primer flank the variant site, whereby for each primer pair, if the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, the first primer is extended to produce a product retaining the label, thus generating a differentially labeled product for each of the allelic forms of the target nucleic acid; and

(b) detecting the labeled product(s), the identity of the label indicating the allelic form(s) of the target nucleic acid present in the sample.

31. The method according to claim 30, wherein the second primers in the different primer pairs are the same.

32. The method according to claim 30, wherein

(i) the sample is from a diploid subject and potentially contains two allelic forms of the target nucleic acid,

(ii) the plurality of primer pairs comprise two primer pairs, the first primer from each primer pair complementary to a different one of the two nucleotides potentially occupying the variant site of the allelic forms, and

(iii) the different labels borne by the different first primers being fluorescent dyes that emit at different wavelengths, whereby detection of emission at two different wavelengths indicates that the subject is a heterozygote, whereas detection of emission at a single wavelength indicates that the subject is a homozygote.

33. The method according to claim 30, further comprises cleaving non-extended first primers with an exonuclease and detecting differentially labeled product(s) by fluorescence polarization.

34. A method for analyzing variant sites in target nucleic acids, comprising:

(a) conducting a sequence-specific amplification step, the amplification step comprising amplifying one or more target nucleic acids by extending a plurality of primer pairs in the presence of a polymerase having a 3'-5' exonuclease activity, each primer pair comprising a first and second primer, wherein

(i) each first primer anneals to a segment that spans a variable site in a target nucleic acid and bears a label on at least one nucleotide other than the 3' terminal nucleotide,

(ii) each second primer is complementary to a segment of a second strand of the target nucleic acid, and the first and second primer flank the variant site; and

(iii) different primer pairs span different variant sites so that the first and second primers for each of the different primer pairs flank different variant sites;

whereby for each primer pair

if the first primer is complementary to the base occupying the variant site in the first strand of the target nucleic acid, the first primer is extended to produce a product retaining the label, and

if the first primer lacks complementarity to the base occupying the variable site in the first strand of the target nucleic acid, the 3'-5' exonuclease activity of the polymerase digests bases from the 3'-end of the first primer, thereby removing the label from the first primer, and wherein the first primer is thereafter extended to produce a product lacking the label,

thus forming a plurality of products from the different primer pairs; and

(b) analyzing the products, the analyzing step comprising determining whether the products retain label to identify the nucleotide at the different variant sites.

35. The method of claim 34, wherein the first primers for different primer pairs are differentially labeled and the analyzing step further comprises detecting the different labels to correlate labeled products with the different variant sites.

36. The method according to claim 35, wherein the labels for the different first primers are fluorescent dyes that emit at different wavelengths.

37. The method of claim 34, wherein the first primers for different primer pairs bear the same label and first and/or second primers from each primer pair are selected to have sequences such that the primers generate different sized extension products during amplification so that products formed from extension of the different primer pairs can be distinguished according to size.

38. The method according to claim 34, wherein the different variant sites are different sites on the single nucleic acid.

39. The method according to claim 34, wherein the different variant sites are different sites on different target nucleic acids.

40. A method for analyzing a variant site in a target nucleic acid, comprising:

(a) conducting a sequence-specific primer extension step, the extension step comprising extending at least one primer that anneals to a segment of the target nucleic acid that spans the variable site and bears a label on at least one nucleotide other than the 3'-terminal nucleotide in the presence of a polymerase having a 3'-5' exonuclease activity, whereby

if the primer is complementary to the base occupying the variable site in the target nucleic acid, the primer is extended to produce a product retaining the label and

if the primer lacks complementarity to the base occupying the variable site in the target nucleic acid, the 3'-5' exonuclease activity of the polymerase digests bases from the 3'-end of the primer, thereby removing the label from the primer, and the primer is thereafter extended to produce a product lacking the label; and

(b) analyzing the product, the analyzing step comprising determining whether the product bears the label to identify the nucleotide at the variant site.

41. The method of claim 40, wherein the extension step is conducted under thermal cycling conditions to linearly amplify the labeled and/or unlabeled product.

42. The method according to claim 40, wherein the labeled nucleotide is aligned with the variable site.

43. The method according to claim 40, wherein the labeled nucleotide is 3' to the nucleotide aligned with the variable site.

44. The method according to claim 42, wherein the nucleotide positions of the primer are numbered consecutively from the 3' end with the 3' terminal nucleotide being designated the first position base, and wherein the labeled nucleotide is at the second, third, fourth or fifth base position.

45. The method according to claim 40, wherein the label is selected from the group consisting of a fluorophore, a chromophore, a radioisotope, an enzyme substrate, an electron dense agent, a magnetic particle, a mass label and an electrochemically active agent.

46. The method according to claim 45, wherein the label is a fluorescent dye selected from the group consisting of FAM, ROX, TAMRA, R110, R6G, Alexa, HEX, TET, Joe, Cy3 and Cy 5.

47. The method according to claim 40, wherein the label is a fluorescent dye and the extension product comprises duplex DNA and the analyzing step comprises:

exposing the extension product to an intercalation dye to introduce a plurality of intercalation dye molecules into the labeled and/or unlabeled extension product; and

detecting fluorescence from the fluorescent dye and the intercalation dye.

48. The method according to claim 47, wherein the fluorescent dye is sufficiently close to the intercalation dye such that energy transfer between the fluorescent dye and the intercalation dye can occur, and wherein the analyzing step further comprises:

exciting the intercalation dye at a first wavelength, whereby fluorescent emission from the intercalation dye is absorbed by the fluorescent dye; and

detecting fluorescence from the fluorescent dye.

49. The method according to claim 48, wherein the analyzing step further comprises determining the ratio between emissions from the fluorescent dye and the intercalation dye.

50. The method according to claim 40, wherein the analyzing step further comprises separating the product from the labeled primer.

51. The method according to claim 40, wherein the product comprises a segment of duplex DNA and the analyzing step further comprises separating the product into single-stranded DNA and determining whether label is retained in one of the single strands.

52. The method according to claim 40, wherein the analyzing step further comprises separating the product from the primer by a size based separation technique.

53. The method according to claim 52, wherein the separation technique is HPLC or electrophoresis.

54. The method according to claim 40, wherein the target nucleic acid is from a diploid subject and potentially comprises a plurality of nucleic acids in which the nucleotide at the variant site differs, whereby if the sample contains the plurality of nucleic acids the primer generates labeled and unlabeled product, and wherein

the detection of only labeled or unlabeled product indicates that the subject is a homozygote, and

the detection of labeled and unlabeled product indicates that the subject is a heterozygote.

55. The method according to claim 54, wherein the label is a fluorescent dye and the analyzing step further comprises:

exposing the product(s) to an intercalation dye to introduce a plurality of intercalation dye molecules into the labeled and unlabeled product; and

detecting fluorescence from the intercalation dye and label wherein

the detection of only product bearing the label and the intercalation dye or only product bearing just intercalation dye indicates that the subject is a homozygote, and

the detection of product bearing both the label and the intercalation dye and product bearing just intercalation dye indicates that the subject is a heterozygote.

56. The method according to claim 40, wherein

(i) the primer bears a second label on a nucleotide other than the one to which the label is attached, and wherein

if the first primer is complementary to the base occupying the variant site in the target nucleic acid, then the product formed retains both labels, and

if the first primer lacks complementarity to the base occupying the variant site in the target nucleic acid, only the second label is retained; and

(ii) the analyzing step further comprises detecting product that retains the label and/or second label.

57. A method for analyzing a variant site in a target nucleic acid, comprising:

(a) conducting a sequence-specific amplification step, the amplification step comprising amplifying the target nucleic acid in a sample containing multiple copies of the target nucleic acid by extending a plurality of primers in the presence of a polymerase having a 3'-5' exonuclease activity, wherein each of the primers anneals to a segment that spans the variable site in the target nucleic acid and bears a label on at least one nucleotide other than the 3' terminal nucleotide, and the primers being selected to be complementary to each of the allelic forms of the target nucleic acid such that different primers are complementary to a different one of the plurality of bases that potentially occupy the variant site and bear different labels; whereby for each primer

if the primer is complementary to the base occupying the variant site in the target nucleic acid, the primer is extended to produce a product retaining the label, thus generating a differentially labeled product for each of the allelic forms of the target nucleic acid; and

(b) detecting the labeled product(s), the identity of the label indicating the allelic form(s) of the target nucleic acid present in the sample.

58. The method according to claim 57, wherein

the sample is from a diploid subject and potentially contains two allelic forms of the target nucleic acid,

the plurality of primers comprise two primers, the two primers complementary to a different one of the two nucleotides potentially occupying the variant site of the allelic forms, and

the different labels borne by the different primers being fluorescent dyes that emit at different wavelengths, whereby detection of emission at two different wavelengths indicates that the subject is a heterozygote, whereas detection of emission at a single wavelength indicates that the subject is a homozygote.

59. A method for analyzing variant sites in target nucleic acids, comprising:

(a) conducting a sequence-specific amplification step, the amplification step comprising amplifying one or more target nucleic acids by extending a plurality of primers in the presence of a polymerase having a 3'-5' exonuclease activity, wherein each primer has a different sequence and anneals to a segment that spans a different variable site in a target nucleic acid and bears a label on at least one nucleotide other than the 3' terminal nucleotide,

whereby for each of the different primers

if the primer is complementary to the base occupying the variant site in the target nucleic acid, the first primer is extended to produce a product retaining the label, and

if the primer lacks complementarity to the base occupying the variable site in the target nucleic acid, the 3'-5' exonuclease activity of the polymerase digests bases from the 3'-end of the first primer, thereby removing the label from the first primer, and wherein the first primer is thereafter extended to produce a product lacking the label,

thus forming a plurality of products; and

(b) analyzing the products, the analyzing step comprising determining whether the products retain label to identify the nucleotide at the different variant sites.

60. The method of claim 59, wherein primers that anneal to different variant sites are differentially labeled and the analyzing step further comprises detecting the different labels to correlate labeled products with the different variant sites.

61. The method of claim 60, wherein labels for the different first primers are fluorescent dyes that emit at different wavelength.

62. The method of claim 59, wherein the products formed from different primers are of differing size.

63. The method according to claim 59, wherein the different variant sites are different sites on the single nucleic acid.

64. The method according to claim 59, wherein the different variant sites are different sites on different target nucleic acids.

65. A kit for analyzing target nucleic acids, comprising:

(a) one or more primers, each primer having a different sequence and capable of annealing to a segment that spans a variable site in a strand of a target nucleic acid and bearing a label or having an attachment moiety for attaching a label on at least one nucleotide other than the 3' terminal nucleotide, different primers bearing different labels or attachment moieties; and

(b) a polymerase having 3'-5' exonuclease activity.

66. The kit of claim 65, wherein the one or more primers comprise primers complementary to the different allelic forms of the target nucleic acid, and primers complementary for different allelic forms bear different labels.

67. The kit of claim 65, wherein each primer bears a second label, the second label being attached to a different nucleotide than the label or attachment moiety.

68. The kit of claim 67, wherein each primer bears the label and the second label and the two labels comprise a fluorescent energy transfer pair.

69. A kit for analyzing target nucleic acids, comprising:

(a) one or more primer pairs, each primer pair comprising a first and second primer, wherein

(i) each first primer anneals to a segment that spans a variable site in a first strand of a target nucleic acid and bears a label or an attachment moiety for attaching a label on at least one nucleotide other than the 3' terminal nucleotide, first primers from different pairs having a different sequence and bearing different labels or attachment moieties,

(ii) each second primer is complementary to a segment of a second strand of the target nucleic acid downstream of the variant site; and

(b) a polymerase having 3'-5' exonuclease activity.

70. The kit of claim 69, wherein the first primers from different primer pairs being complementary to different allelic forms of the same variant site on the target nucleic acid(s), and the second primer of each primer pair having the same sequence.

71. The kit of claim 69, wherein the different primer pairs are selected to amplify different variant sites.

72. The kit of claim 71, wherein each second primer bears a label.

73. A kit for analyzing a variant site in a target nucleic acid, comprising:

(a) one or more primers, each primer comprising a modified nucleotide for attachment of label and capable of annealing to a segment that spans the variable site in a strand of the target nucleic acid, the modified nucleotide being other than the 3' nucleotide;

(b) a label that can be attached to the first primer; and

(c) a polymerase having 3'-5' exonuclease activity.

74. The kit of claim 73, further comprising an intercalation dye.

75. The kit of claim 73, further comprising a second primer for each primer, each second primer complementary to a segment of a second strand of the target nucleic acid.

76. The kit of claim 74, wherein the label is a fluorescent dye that fluoresces at a longer wavelength than the intercalation dye.

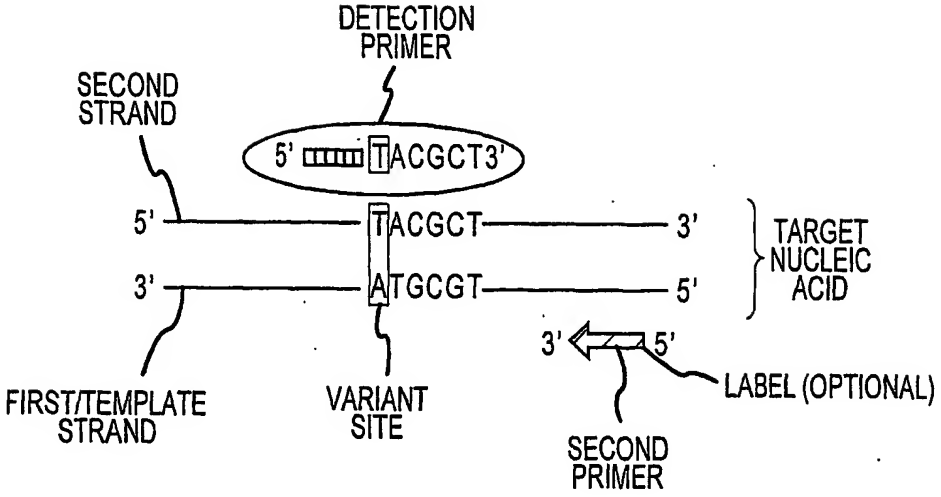

DETECTION PRIMER
SECOND STRAND
5' TACGCT3'
5'
TACGCT
3'
3'
ATGCGT
5'
TARGET NUCLEIC ACID
3'
5'
LABEL (OPTIONAL)
FIRST/TEMPLATE STRAND
VARIANT SITE
SECOND PRIMER

FIG.1A

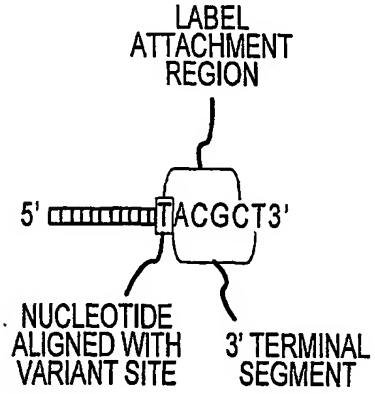

LABEL ATTACHMENT REGION
5' TACGCT3'
NUCLEOTIDE ALIGNED WITH VARIANT SITE
3' TERMINAL SEGMENT

FIG.1B

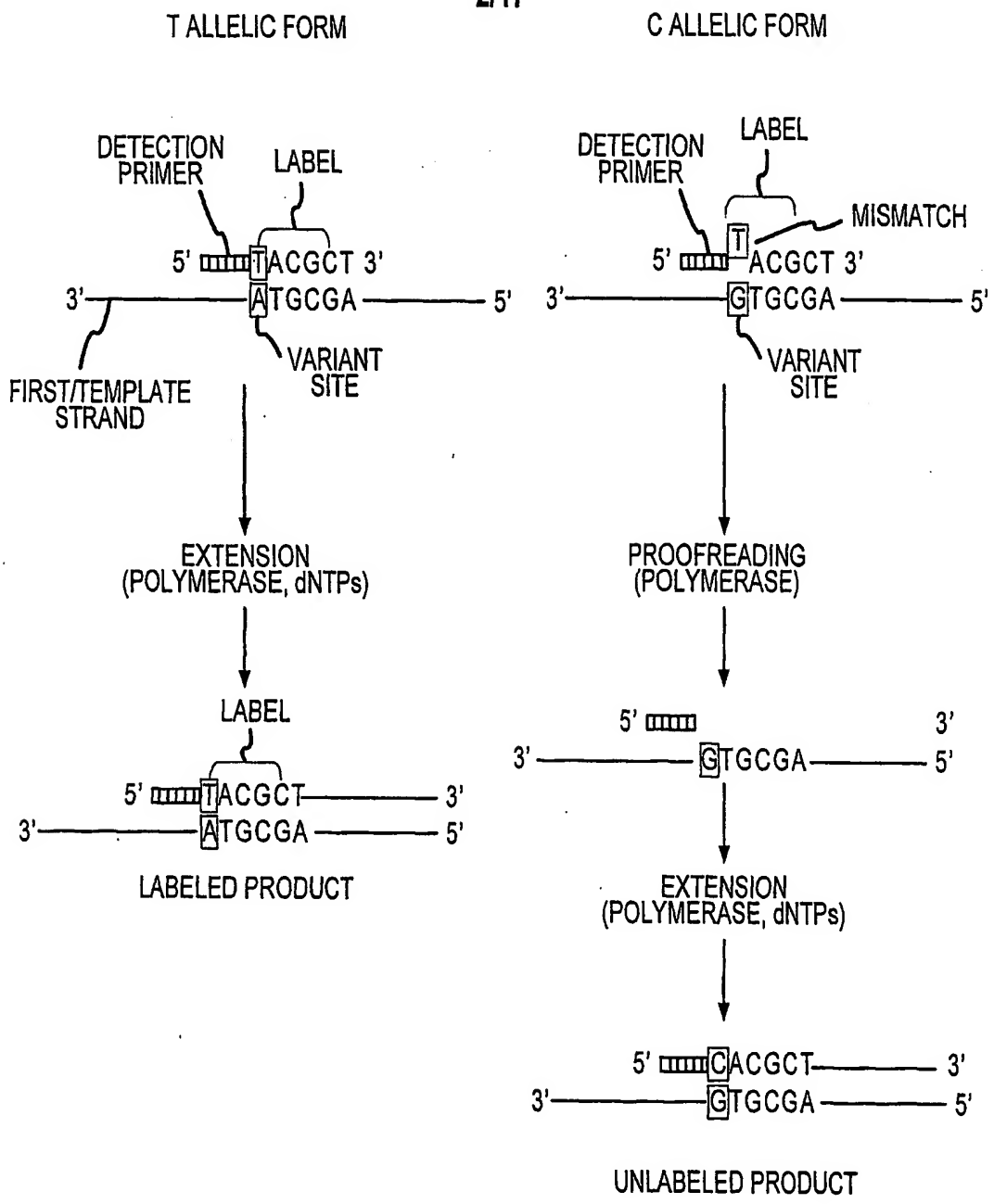

T ALLELIC FORM
C ALLELIC FORM
DETECTION PRIMER
LABEL
5' TACGCT 3'
3' ATGCGA 5'
VARIANT SITE
FIRST/TEMPLATE STRAND
EXTENSION (POLYMERASE, dNTPs)
LABEL
5' TACGCT 3'
3' ATGCGA 5'
LABELED PRODUCT
DETECTION PRIMER
LABEL
MISMATCH
5' TACGCT 3'
3' GTGCGA 5'
VARIANT SITE
PROOFREADING (POLYMERASE)
5' 3'
3' GTGCGA 5'
EXTENSION (POLYMERASE, dNTPs)
5' CACGCT 3'
3' GTGCGA 5'
UNLABELED PRODUCT

FIG.1C

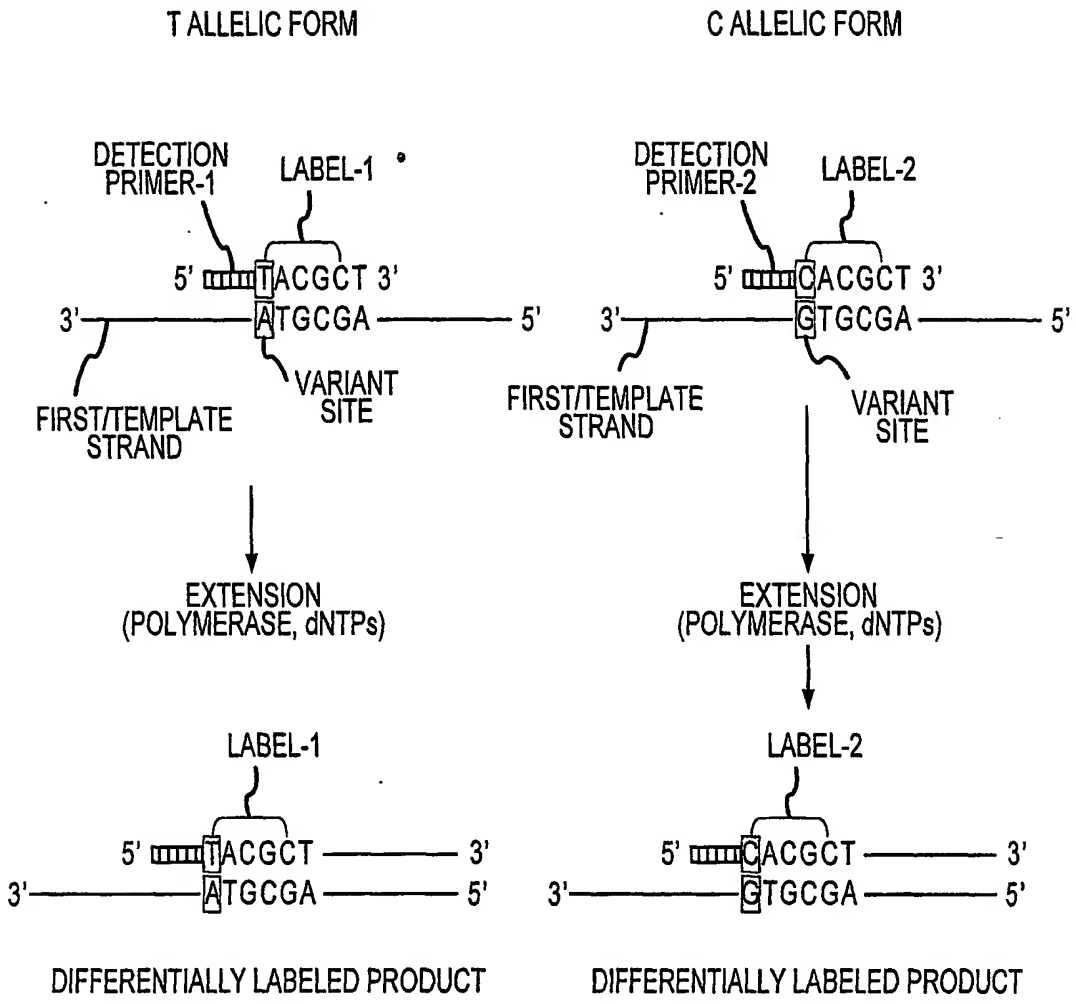

T ALLELIC FORM
C ALLELIC FORM
DETECTION PRIMER-1
LABEL-1
5' TACGCT 3'
3' ATGCGA 5'
FIRST/TEMPLATE STRAND
VARIANT SITE
EXTENSION (POLYMERASE, dNTPs)
LABEL-1
5' TACGCT 3'
3' ATGCGA 5'
DIFFERENTIALLY LABELED PRODUCT
DETECTION PRIMER-2
LABEL-2
5' CACGCT 3'
3' GTGCGA 5'
FIRST/TEMPLATE STRAND
VARIANT SITE
EXTENSION (POLYMERASE, dNTPs)
LABEL-2
5' CACGCT 3'
3' GTGCGA 5'
DIFFERENTIALLY LABELED PRODUCT

FIG.1D

8341 ccatgtgtac ccataaaatg aattacacag agatcgctat aggatttaaa gcktttatac
8401 taaatgtgct gggattttgc aaactatagt gtgctgttat tgttaattta aaaaaactct
8461 aagttaggat tgacaaatta tttctctttа gtcatttgct tgtatcacca aagaagcaaa

FIG.2A

FAM

5′ ag agatcgctat aggatttaaa gcTt 3′

FIG.2B

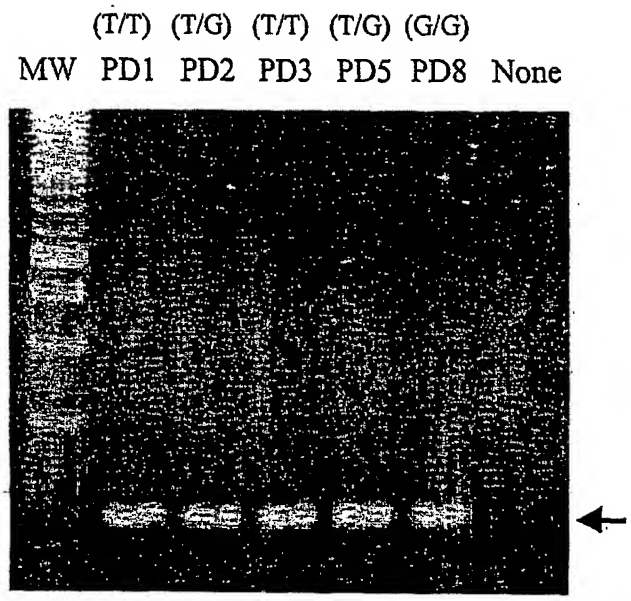
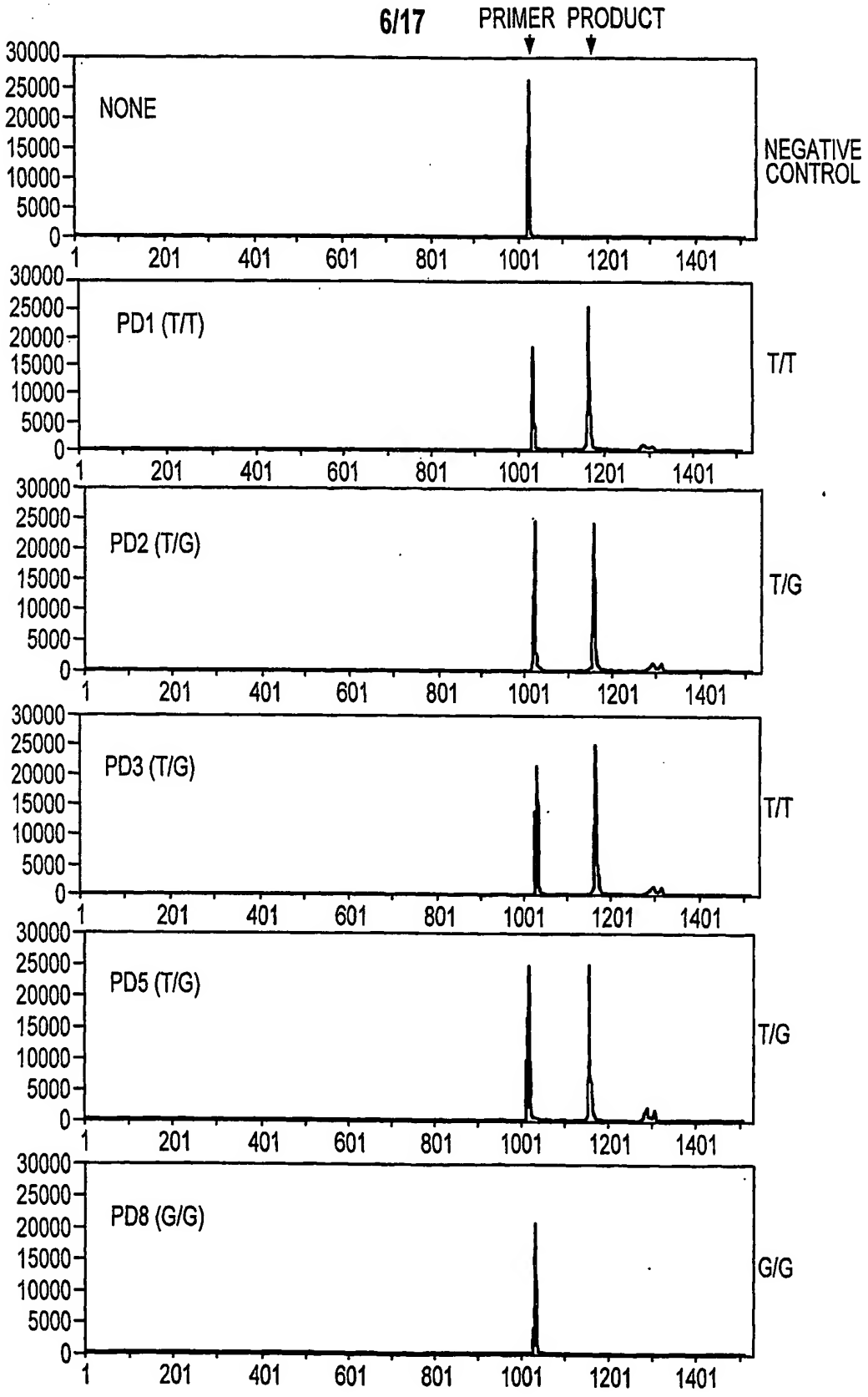


FIG.2C

6/17


PRIMER PRODUCT
NONE
NEGATIVE CONTROL
PD1 (T/T)
T/T
PD2 (T/G)
T/G
PD3 (T/G)
T/T
PD5 (T/G)
T/G
PD8 (G/G)
G/G
30000
25000
20000
15000
10000
5000
0
1
201
401
601
801
1001
1201
1401

```
5521 aataagttct atgatgcacc tactagacac ctamtctgca ctagatggtg ggggaattaa
5581 gagcatgggc atgatcctgt gaccggaagc ccgcttacag tcagggtgga ggacagacct
5641 actcatgaaa caaacacagt gacatatagt gacacagaag caaatgtcaa atatgcttgc
```

FIG.3A

FAM

5′ atgatgcacc tactagacac ctaTt 3′

FIG.3B

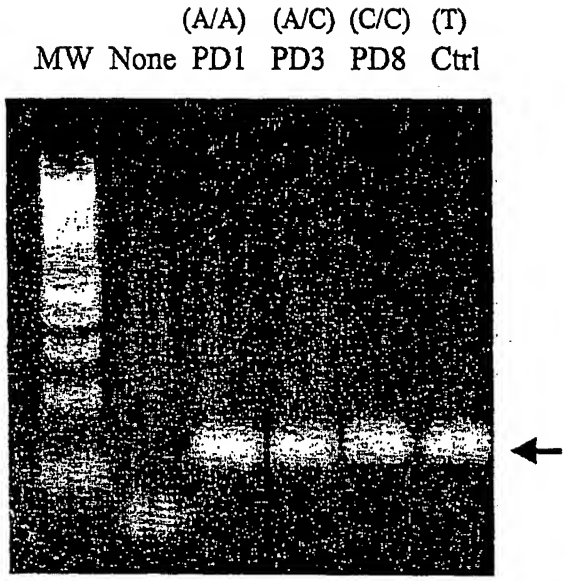


FIG.3C

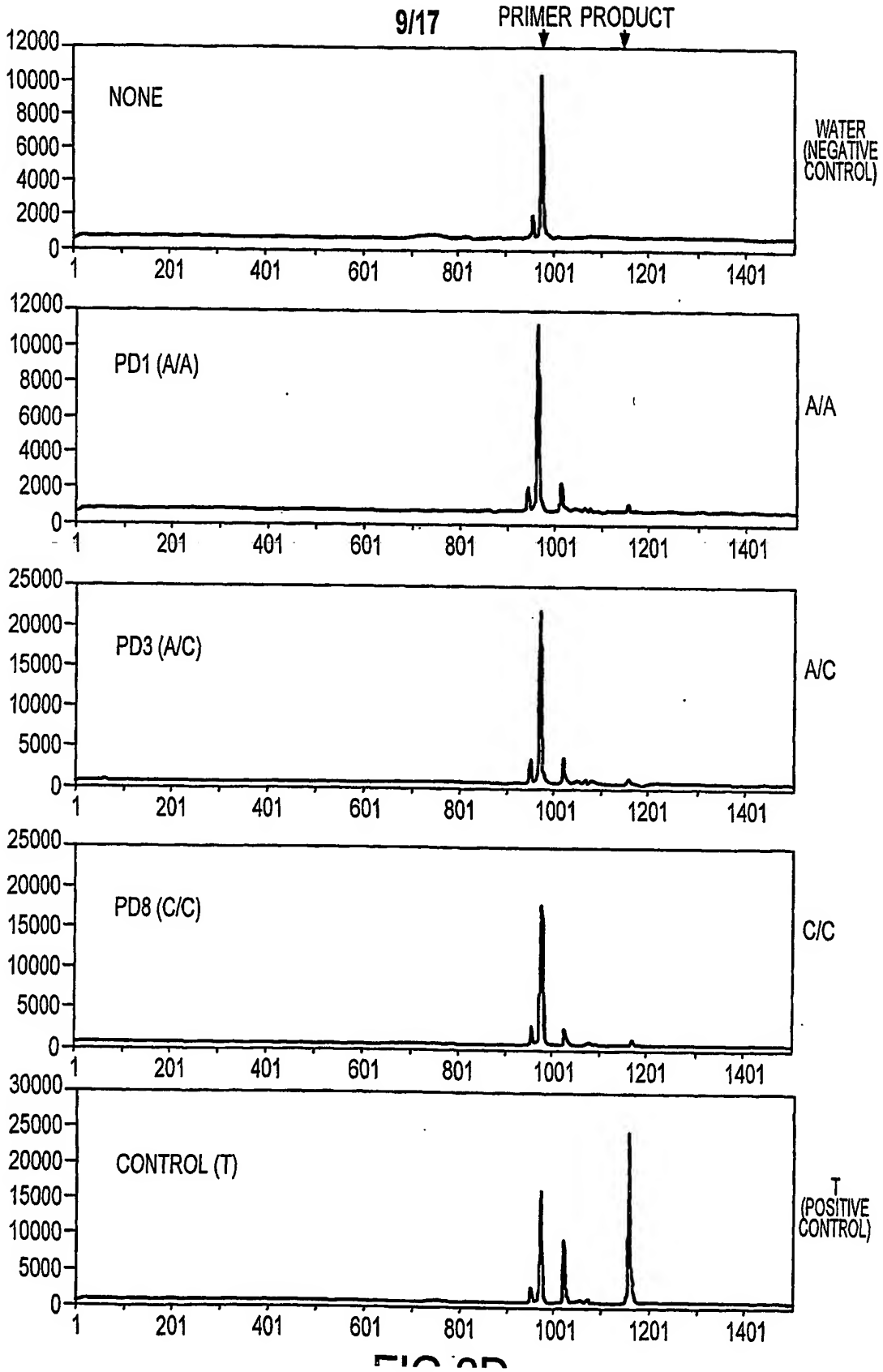

PRIMER PRODUCT
NONE
WATER (NEGATIVE CONTROL)
PD1 (A/A)
A/A
PD3 (A/C)
A/C
PD8 (C/C)
C/C
CONTROL (T)
T (POSITIVE CONTROL)
1
201
401
601
801
1001
1201
1401

FIG 2D

Primer-1:

Rox

5′ atgatgcacc tactagacac ctaTt 3′

Primer-2:

Rox

5′ atgatgcacc tactagacac ctaTtc 3′

FIG.4A

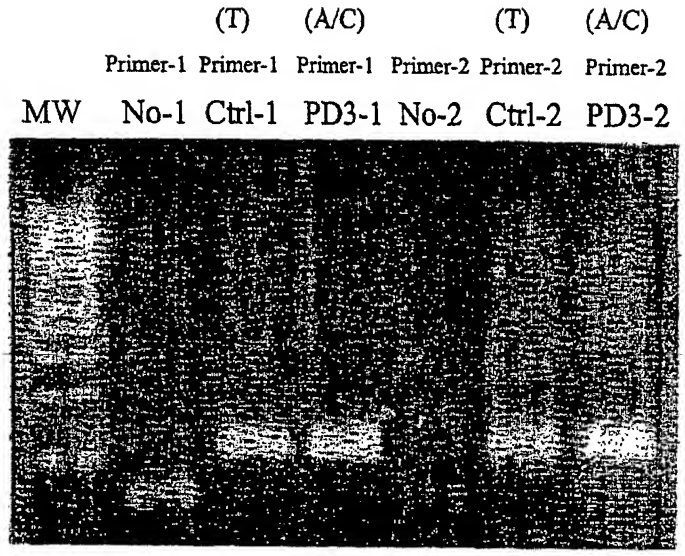

(T) (A/C) (T) (A/C)
Primer-1 Primer-1 Primer-1 Primer-2 Primer-2 Primer-2
MW No-1 Ctrl-1 PD3-1 No-2 Ctrl-2 PD3-2

FIG.4B

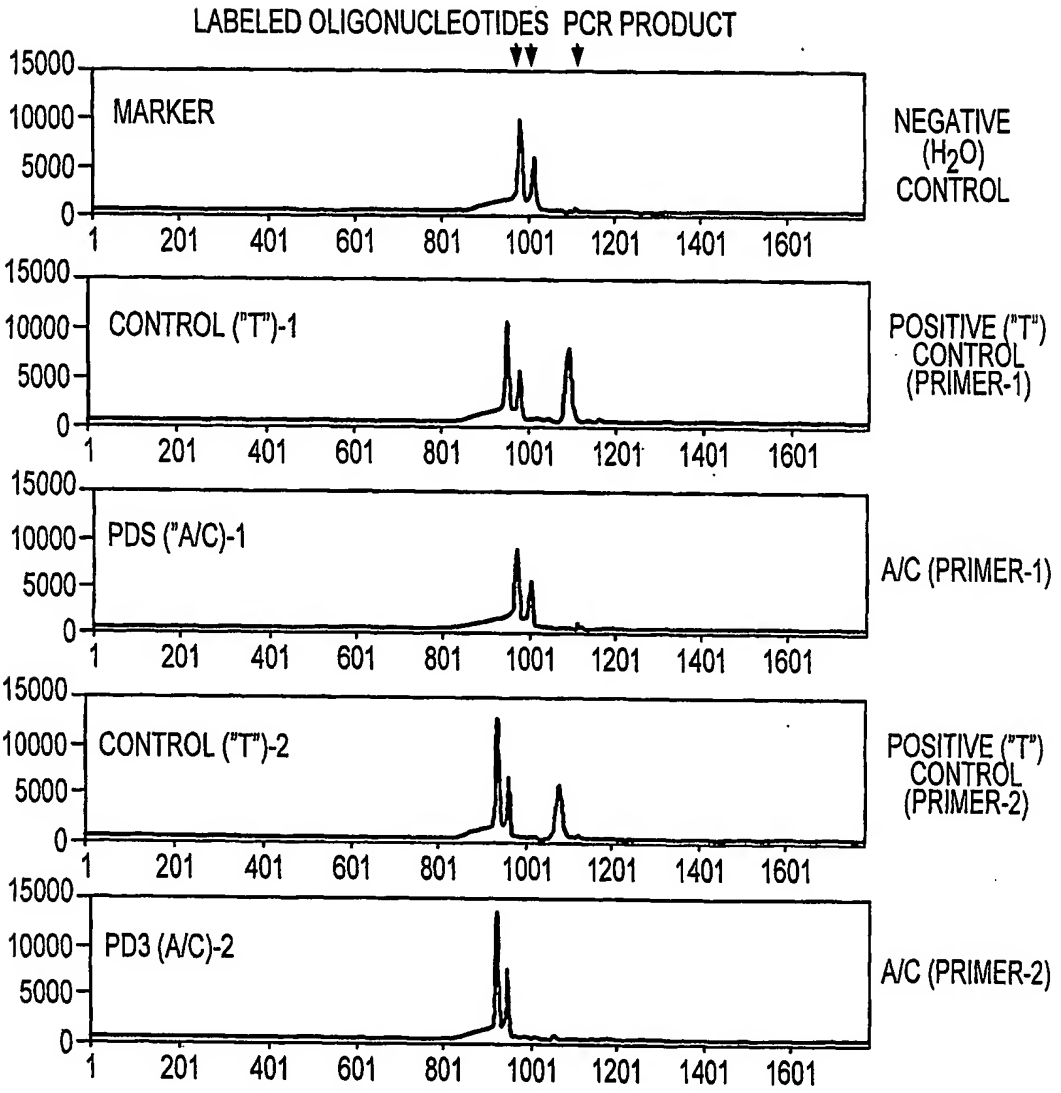
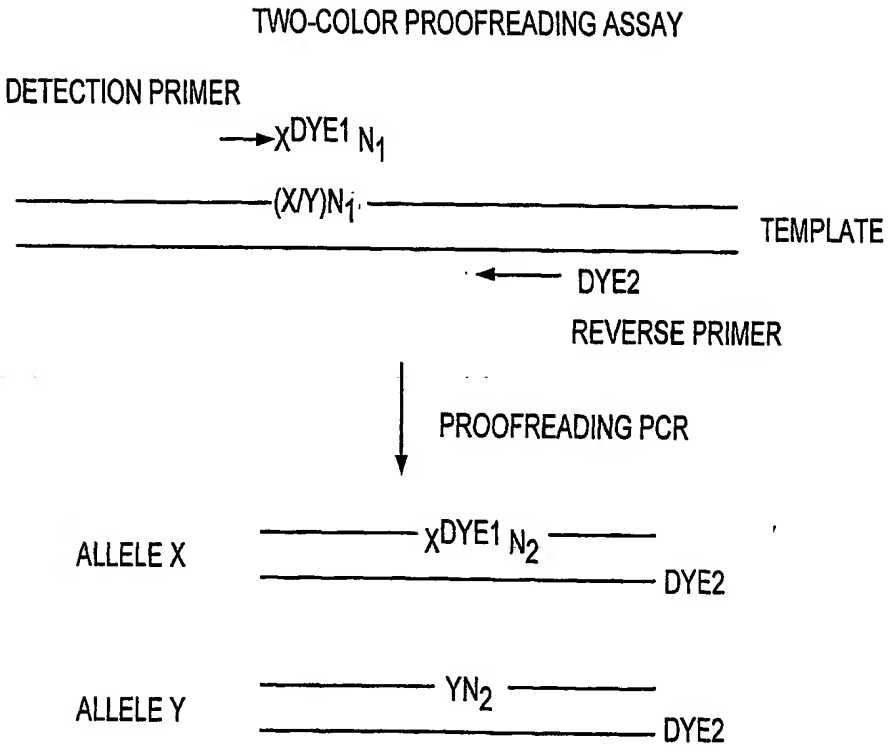

LABELED OLIGONUCLEOTIDES PCR PRODUCT
MARKER
NEGATIVE (H_2O) CONTROL
CONTROL ("T")-1
POSITIVE ("T") CONTROL (PRIMER-1)
PDS ("A/C)-1
A/C (PRIMER-1)
CONTROL ("T")-2
POSITIVE ("T") CONTROL (PRIMER-2)
PD3 (A/C)-2
A/C (PRIMER-2)
15000
10000
5000
0
1
201
401
601
801
1001
1201
1401
1601

FIG.4C



GENOTYPES DETERMINED BY THE RATIO OF THE TWO DYES.

FIG.5

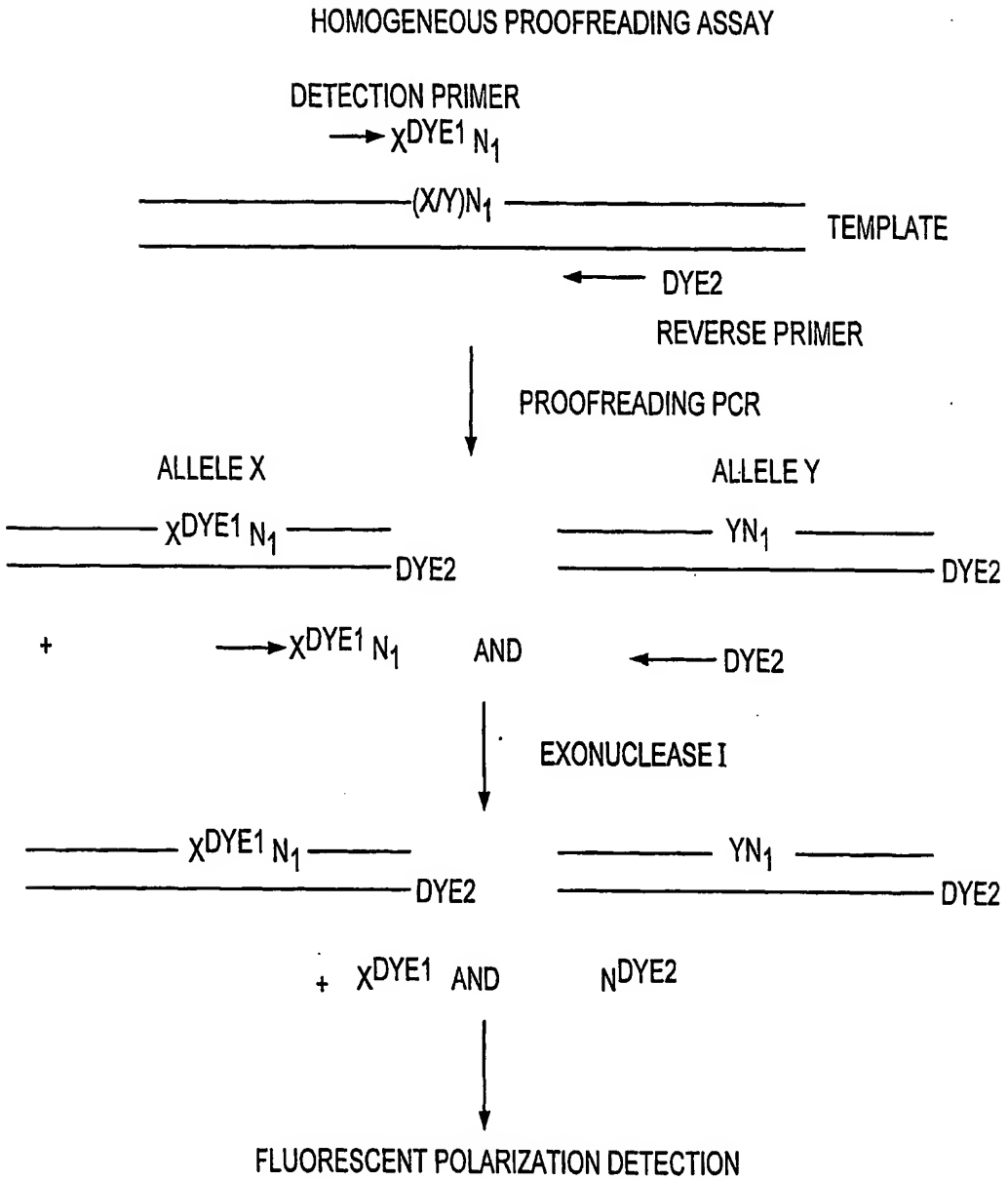

HOMOGENEOUS PROOFREADING ASSAY
DETECTION PRIMER
XDYE1 N1
(X/Y)N1
TEMPLATE
DYE2
REVERSE PRIMER
PROOFREADING PCR
ALLELE X
ALLELE Y
XDYE1 N1
DYE2
YN1
DYE2
+
XDYE1 N1
AND
DYE2
EXONUCLEASE I
XDYE1 N1
DYE2
YN1
DYE2
+ XDYE1 AND NDYE2
FLUORESCENT POLARIZATION DETECTION

FIG.6

TWO COLOR PROOFREADING ASSAY
AT THE LPL-5 SNP SITE.

SEQUENCE OF THE LPL-5 SNP:

ctgtgggacc ataatcttga agwcacagac aggcttcact
catccctgcc tcctgcacca gtgggttcaa ggctctgtca
gtgtccccta ggggcacctc accactccca gcttcttcag
ctctggcctg tcctgctgcc tgcaagggtt ttgcttaatt
ctcaattcaa tgtctcttca

FIG.7A

DETECTION PRIMER FOR LPL-5: R110

5′ tgt ggg acc ata atc ttg aag Tc 3′

FIG.7B

REVERSE PRIMER FOR LPL-5:

5′ ROX— acc ctt gca ggc agc a 3′

FIG.7C

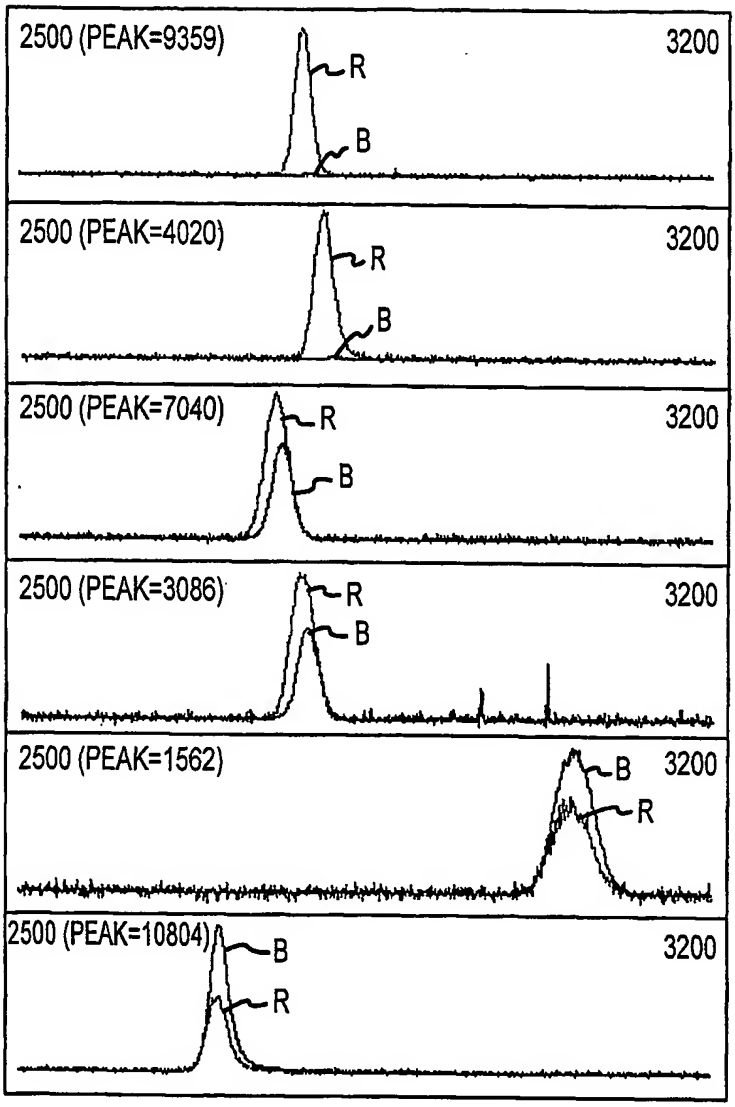

2500 (PEAK=9359)
3200
R
B
PDR02 (A/A)
2500 (PEAK=4020)
3200
R
B
PDR05 (A/A)
2500 (PEAK=7040)
3200
R
B
PDR03 (A/T)
2500 (PEAK=3086)
3200
R
B
PDR06 (A/T)
2500 (PEAK=1562)
3200
B
R
PDR08 (T/T)
2500 (PEAK=10804)
3200
B
R
HD1.10 (T/T)

FIG.8

17/17

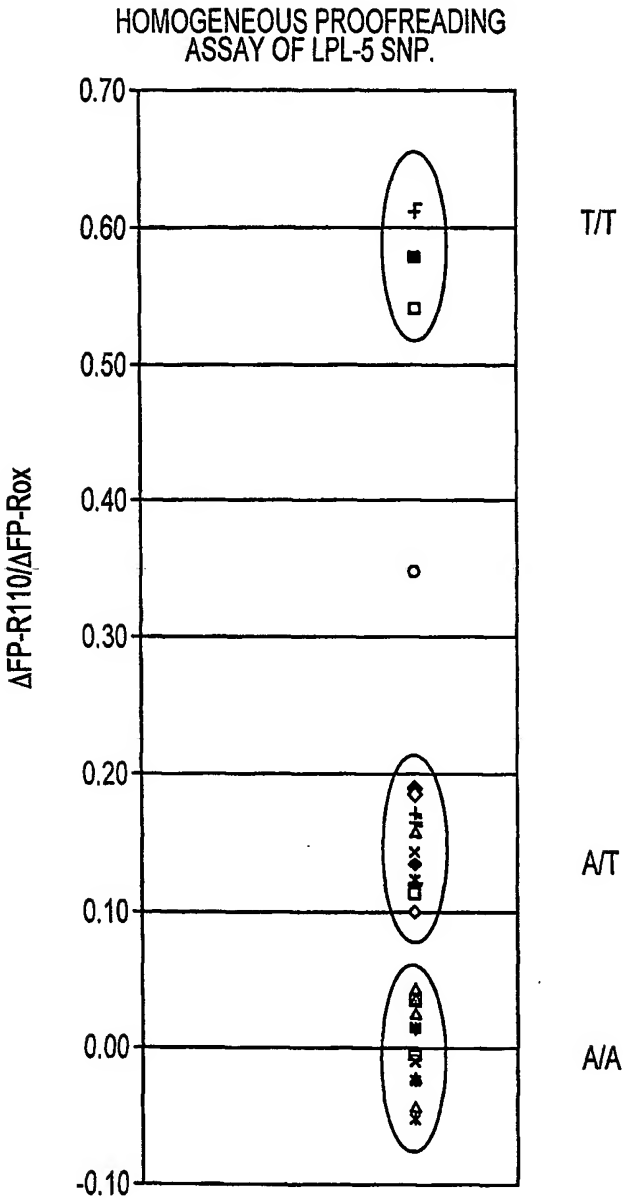

HOMOGENEOUS PROOFREADING
ASSAY OF LPL-5 SNP.
ΔFP-R110/ΔFP-Rox
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
T/T
A/T
A/A

FIG.9